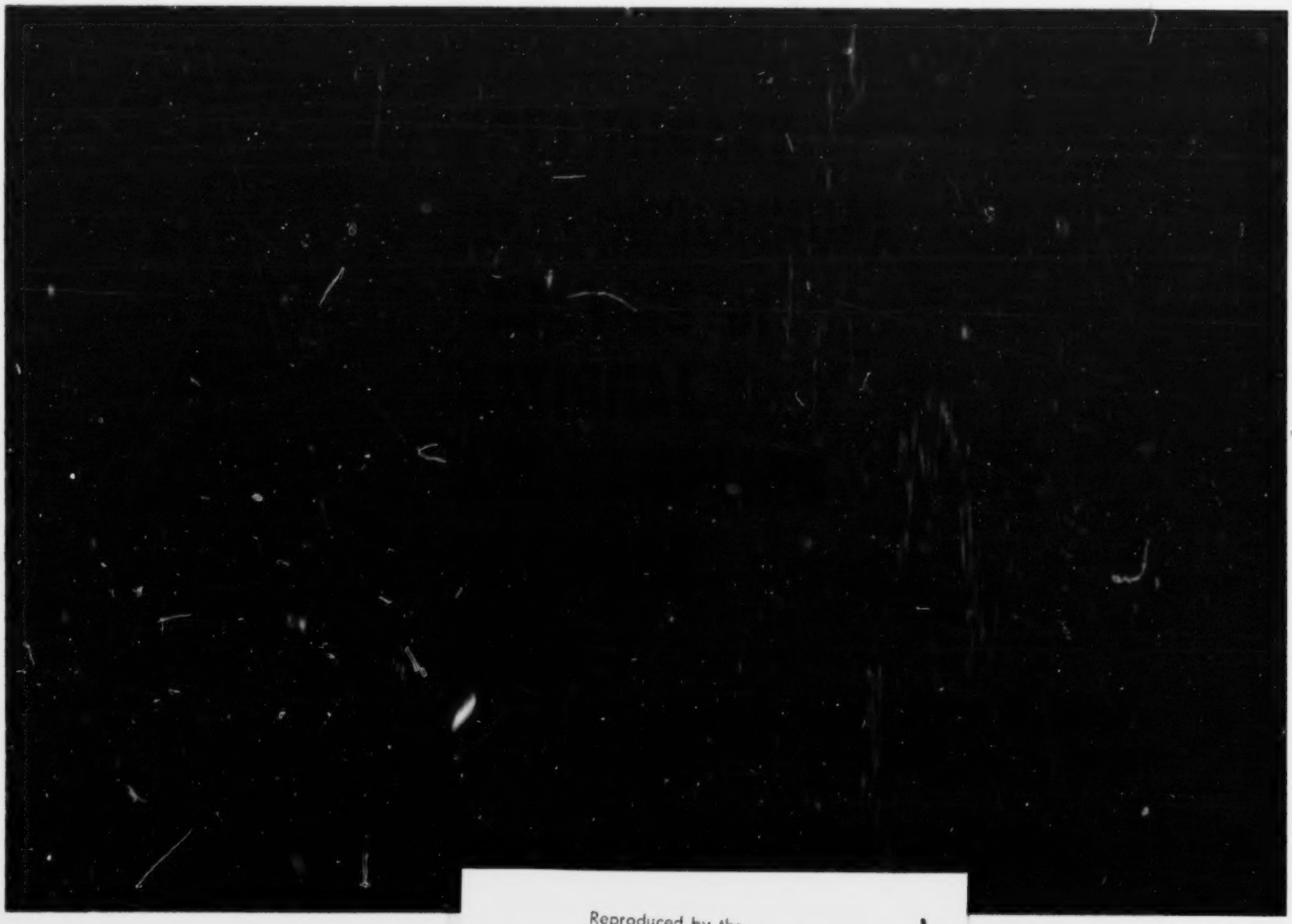
M(051)
T779tr
no.19
c.2

A Cooperative Program with the U. S. Weather Bureau and the U. S. Public Health Service, under Contract Cwb 10014.

PB176526



Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

THE TRAVELERS RESEARCH CENTER INC.
650 MAIN STREET HARTFORD 3 CONNECTICUT

Contract Cwb 10014

- FINAL REPORT -

# PRELIMINARY METEOROLOGICAL ANALYSIS

# OF

# NATIONAL AIR SAMPLING NETWORK DATA

by Joseph G. Bryan
Glenn R. Hilst

LIBRARY
JUL 12 1963
WEATHER BUREAU

[Travelers Weather Research Center, Inc.,
Hartford, Conn.]

TRC-19
January 1962

Volume II:
Comparison and Interpretation
of Results

124 393

M(051)
T779tr
No.19
c.2

## PREFACE

This Report is the second of a two-volume presentation of the analysis of relationships between particulate pollution levels and concomitant meteorological variables, both measured in urban areas. Data from 51 stations for the period 1957 through 1959 have been used for these analyses.

The methodology for statistical discernment of the relationships between urban pollution levels and the meteorological variables observed routinely at nearby U. S. Weather Bureau stations for each area has been presented in Volume I. The results of application of these analytical techniques have also been presented there. In the present volume the comparison and interpretation of these results are presented and suggestions for future studies are outlined.

TABLE OF CONTENTS


| | | Page |
|---|---|---|
| | Title Page | |
| | Preface | i |
| | Table of Contents | iii |
| | List of Tables | v |
| | List of Figures | vii |
| I. | Introduction | 1 |
| II. | Particulate Pollution Experience in the United States (1957-1959) | 3 |
| | A. The Dependence of $\bar{L}$ and $\sigma_L$ upon Metropolitan Populations | 10 |
| | B. The Dependence of $\bar{L}$ and $\sigma_L$ upon Geographical Location | 15 |
| III. | The Predictability of Particulate Pollution Levels from Conventional Meteorological Measurements | 25 |
| | A. Predictability | 26 |
| | B. The Contributions of Seasonal Meteorological Variability and Day-of-the-Week to Two-Group Pollution Prediction | 27 |
| | C. The Contributions of Concurrent Meteorological Variability to Two-Group Pollution Prediction | 37 |
| | D. The Total Predictability of Two-Group Pollution Levels by Presently Available Independent Variables | 41 |
| | E. Total Predictability as a Function of the Frequency Distribution of Concentration | 42 |

TABLE OF CONTENTS (Continued)


| | | |
|---|---|---|
| IV. | Comparison of the Predictive Information contained in the NASN Meteorological Variables | 49 |
| | A. The Frequency of Selection of each Meteorological Predictor and its Average Contribution to Two-Group Discrimination | 49 |
| | B. The Total Contribution of each Predictor to the Reduction of Mean Square Error of Two-Group Discrimination | 53 |
| | C. In Summary | 57 |
| V. | Tests for Homogeneity Among NASN Station | 59 |
| | Appendix A | |

## LIST OF TABLES


| Table | Title | Page |
|---|---|---|
| 1. | Measures of frequency distributions for NASN Stations (1957-1959) | 8, 9 |
| 2. | Predictor Identification Code | 29 |
| 3. | Background variations in pollution due to day of week and seasonal effects of weather | 31, 32 |
| 4. | Frequency of occurrance of various ranges of $\Delta$MSE(M) at 49 National Air Sampling Network Stations | 40 |
| 5. | Frequency of various ranges of total reduction in mean square of two-group pollution level discrimination chieved by all NASN predictors | 41 |
| 6. | Summary of the frequency of choice of each meteorological variable and the average reduction in mean square error | 51 |
| 7. | Ranking of meteorological predictors on the basis of total contribution to the predictability of two-group pollution levels | 53 |
| 8. | Test of homogeneity criteria for NASN Stations | 62 |

LIST OF FIGURES


| Figure | Description | Page |
|---|---|---|
| 1 | Frequency distribution of concentration of total Particulates at Columbus, Ohio. | 3 |
| 2 | Comparison of joint values of median concentration and average logarithm of concentration for 51 urban stations. Dashed line is $L = \ln \chi_{med}$. | 5 |
| 3 | Joint values of central tendency and dispersion of frequency distributions of $\ln \chi$ as measured by L and $\sigma_L$ for 51 urban stations. Data for period 1957 - 1959. (See Table 1 for station identification, e.g., DCA = Washington, D.C.) | 6 |
| 4 | Joint values of central tendency and dispersion of frequency distribution of $\ln \chi$, as measured by L and $\sigma_L$, for 51 urban stations. Data for period 1957 - 1959. (See Table 1 for station identification, e.g., DCA = Washington, D.C.) | 7 |
| 5 | Average of $\ln \chi$ as function of metropolitan population. Regression line given by $\bar{L} = 4.81 + .17 \ln P_o/10^6$. | 13 |
| 6 | Average of $\ln \chi$ as function of metropolitan population. Regression line given by $\bar{L} = 2.19 + .20 \ln P_o/10^6$. | 14 |
| 7 | Values of average of $\ln$ of total particulate concentrations ($\bar{L}$) observed in the continental United States during 1957 - 1959. (National Air Sampling Network Data.) | 16 |
| 8 | Values of the standard deviation of $\ln$ of total particulate concentrations ($\sigma_L$) observed in continental United States during 1957 - 59. (National Sampling Network Data.) | 17 |
| 9 | Values of average of $\ln$ total particulate concentration ($\bar{L}^*$) normalized to a base metropolitan population of $10^6$, as observed during period 1957 - 1959. (National Air Sampling Network Data.) | 18 |
| 10 | Values of average of $\ln$ of organic particulate concentrations ($\bar{L}$) observed in continental United States during 1957 - 1959. (National Air Sampling Network Data.) | 19 |
| 11 | Values of standard deviation of $\ln$ of organic particulate concentrations observed in continental United States during 1957 - 1959. (National Air Sampling Network Data.) | 20 |

LIST OF FIGURES (Contined)


| | | |
|---|---|---|
| 12 | Values of average of $\ln$ of organic particulate concentrations normalized to a base population of $10^6$ as observed during period 1957 - 1959. (National Air Sampling Network Data.) | 21 |
| 13 | Reduction of mean square error (MSE) of discrimination between High and Low values of total and organic particulates for each predictor selected by two-group MDA. Predictors listed at bottom of figures were selected in order shown. (See Table 2 for identification of predictors.) | 28 |
| 14 | The relative contributions of seasonal meteorological variability (ΔMSE (W)) and the day-of-the week (ΔMSE (D)) to the discrimination of HIGH and LOW total particulate pollution levels. | 33 |
| 15 | The relative contribution of seasonal meteorological variability (ΔMSE (W)) and the day-of-the-week (ΔMSE(D)) to the discrimination of HIGH and LOW organic particulate pollution levels. | 34 |
| 16 | The reduction of the mean square error of choice between HIGH and LOW pollution groups attributable to the seasonal weather and day-of-the-week variations, ΔMSE (SB). | 35 |
| 17 | The reduction of the mean square error of choice between HIGH and LOW pollution groups attributable to the seasonal weather and day-of-the-week variations, ΔMSE (SB). | 36 |
| 18 | The reduction of mean square error of discrimination of HIGH and LOW pollution levels attributable to concurrent meteorological variables (ΔMSE (M)). | 38 |
| 19 | The reduction of mean square error of discrimination of HIGH and LOW pollution levels attributable to concurrent meteorological variables (ΔMSE (M)). | 39 |
| 20 | Total reduction ir mean square error of discrimination of HIGH and LOW pollution levels achieved by all predictors available from National Air Sampling Network Data. | 43 |
| 21 | Total reduction in mean square error of discrimination of HIGH and LOW pollution levels achieved by all predictors available from National Air Sampling Network Data. | 44 |

LIST OF FIGURES (Continued)


| | | |
|---|---|---|
| 22 | Minimum MSE as related to central tendency and dispersion of frequency distributions of $\ln X$ for 51 urban stations. Data for period 1957 - 1959. (Individual stations may be identified by reference to Fig. 3.) | 46 |
| 23 | Minimum MSE as related to central tendency and dispersion of frequency distribution of $\ln X$ for 51 urban stations. Data for period 1957 - 1959. (Individual stations may be identified by reference to Fig. 4.) | 47 |

# I INTRODUCTION

Methods of analysis of concomitant observations of particulate pollution levels and local meteorological conditions which are designed to discern relationships between the pollution level and various meteorologic quantities have been presented in Volume I of this Final Report. From the methods tested, two-group discriminant analysis was selected for the initial study and data from forty-nine stations within the continental United States, and one each from Alaska and Hawaii were subjected to this analysis technique. These observations consisted of 24-hour samples taken at bi-weekly intervals (but at random times and days within each two week period) during the three years 1957-1959. The quantities observed are listed in Appendix C of Volume I and are repeated in Appendix A to the present Volume.

The results of the initial analyses in Volume I, combined with various measures of the pollution experience at each station during 1957-1959, form the basis for the present comparison and interpretation of pollution experience and predictability among the 51 cities.

It has been recognized for some time that urban air pollution is dependent upon a complex interaction between source variables and atmospheric transport, diffusion, and deposition processes. Since no measures of the source configurations and strengths could be included in the present study, it has not been possible to separate this contribution to pollution variability. This handicap, coupled with the fact that the meteorological variables chosen as predictors of pollution levels are, for the most part, only secondarily related to atmospheric transport and diffusion (the outstanding exception is wind velocity) presented formidable obstacles to the realization of the objectives of this study. However, despite these handicaps, the results of this study are quite encouraging and valuable, both in that positive and demonstrable results have been obtained, and guidelines for further research and experimentation can be delineated.

## II. PARTICULATE POLLUTION EXPERIENCE IN THE UNITED STATES (1957-1959)

The principal measures of particulate air pollution levels available from the National Air Sampling Network have been described and presented in Volume I of this report. The data consist of 24-hour averages of the airborne concentrations of total and organic particulates, sampled once each two weeks near the center of the city. (The actual three-year sample size for each city included in the present study ranges from 46 to 78).

These samples of 24-hour average concentrations for each city permit the definition of a frequency distribution of particulate pollution levels at the sampling site. A typical distribution is shown in Figure 1, where the frequency of occurrence of concentrations of total particulates is given in class intervals of $50 \mu g/m^3$. As shown by Stern[1] these distributions are generally skewed but can be normalized by the transformation $Y = \ln \chi$, where Y is the new variable and $\chi$ is the original measure of concentration. Then, Y is normally distributed with mean $\bar{L}$ and variance $\sigma_L^2$, and these two measures completely specify the distribution of Y, or, by inverse transformation, $\chi$. If Y is normally distributed, $\bar{L} = \ln \chi_{med}$, since the mean of the distribution of Y must coincide with the modal and median value of Y.

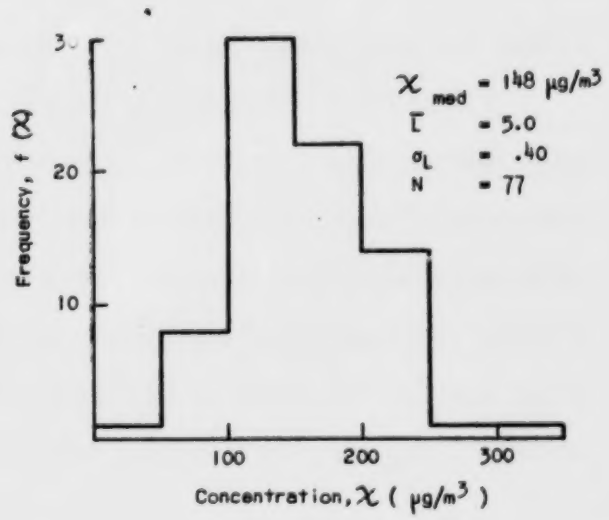



Fig. 1. Frequency distribution of concentration of total particulates at Columbus, Ohio.

---

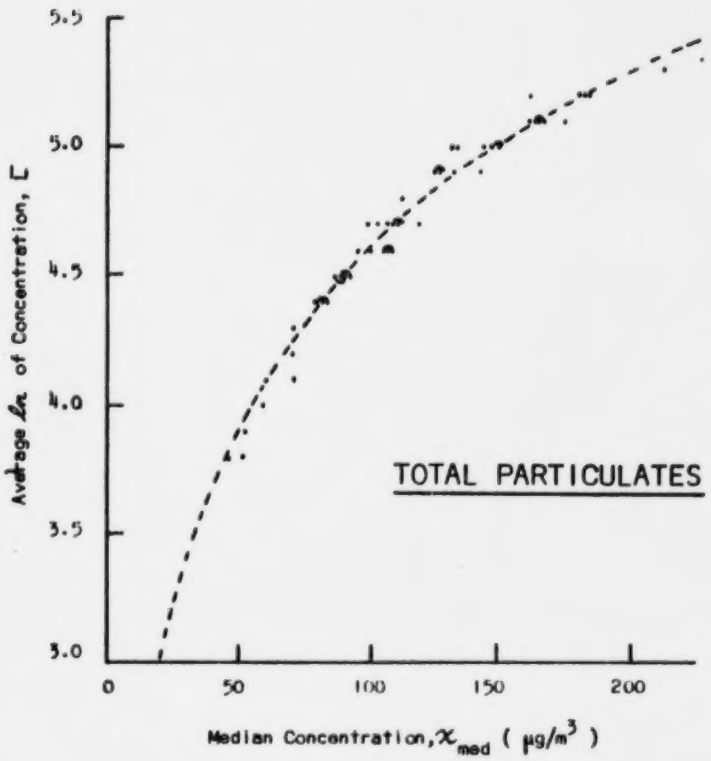
1 See Stern, A. C., 1957. Results of Three Years of Operation of the National Air Sampling Network. J. of Air Poll. Cont. Assoc., Nov. 1957.

A plot of $\bar{L}$ versus $\chi_{med}$ for both total particulates and organic particulates is shown in Figure 2. The general verification of $\bar{L} = \ln \chi_{med}$ is evident; $\bar{L}$ and $\chi_{med}$ may be used interchangeably as the measure of central tendency of the pollution level distributions. On this basis, and for consistency, $\bar{L}$ and $\sigma_L$ have been chosen as the primary measures of central tendency and dispersion, respectively.

While it is not the purpose of this study to explain or predict the median level of concentration observed at various cities, it is of interest to compare the pollution distributions experienced at various stations as background for the interpretation of the predictability of HIGH and LOW values of concentration and for studies of homogeneity of pollution among various cities.

The joint values of $\bar{L}$ and $\sigma_L$ for each station and type of particulate pollution, listed in Table 1, are plotted in Figures 3 and 4 for direct comparison. Among these 51 cities there is a broad range of both $\bar{L}$ and $\sigma_L$, but no evident correlation between these two measures. The distributions of $\ln \chi$ implied by these measures range from consistently high levels (relative to other stations), such as at Indianapolis (IND) for total particulates and at Baltimore (BAL) for organic particulates, to low median values and wide range, such as at Helena (HLN) for total particulates.

These joint values of $\bar{L}$ and $\sigma_L$ will be utilized more extensively in later discussions. However, it is worthy of note at this point that the distributions of total particulates and organic particulates may be significantly different at any given station. This is the first suggestion that the sources of these two classes of pollution may be quite different, a point which will arise again in the study of interrelationship between pollution level (HIGH or LOW) and various meteorological variables.



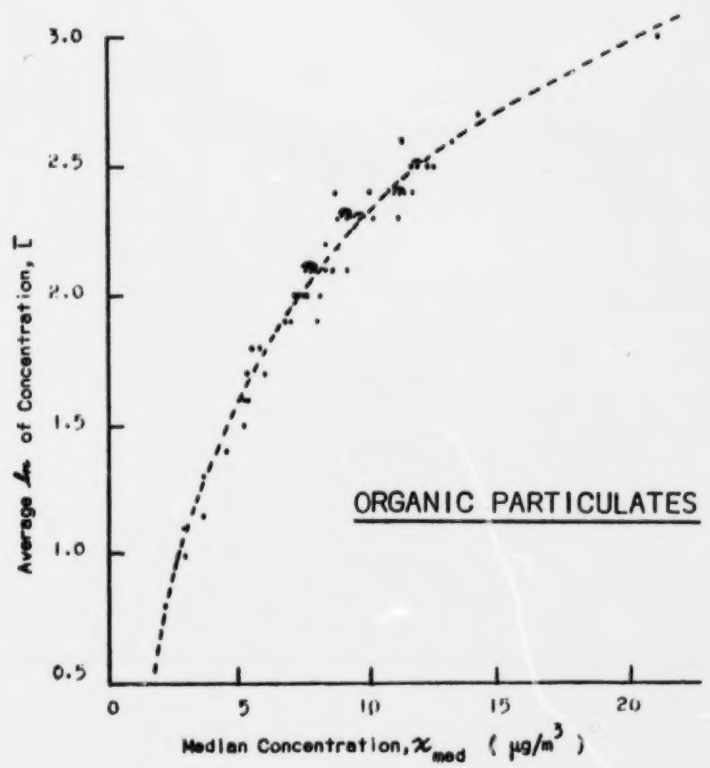




Fig. 2. Comparison of joint values of median concentration and average logarithm of concentration for 51 urban stations. Dashed line is $L = \ln \chi_{med}$.

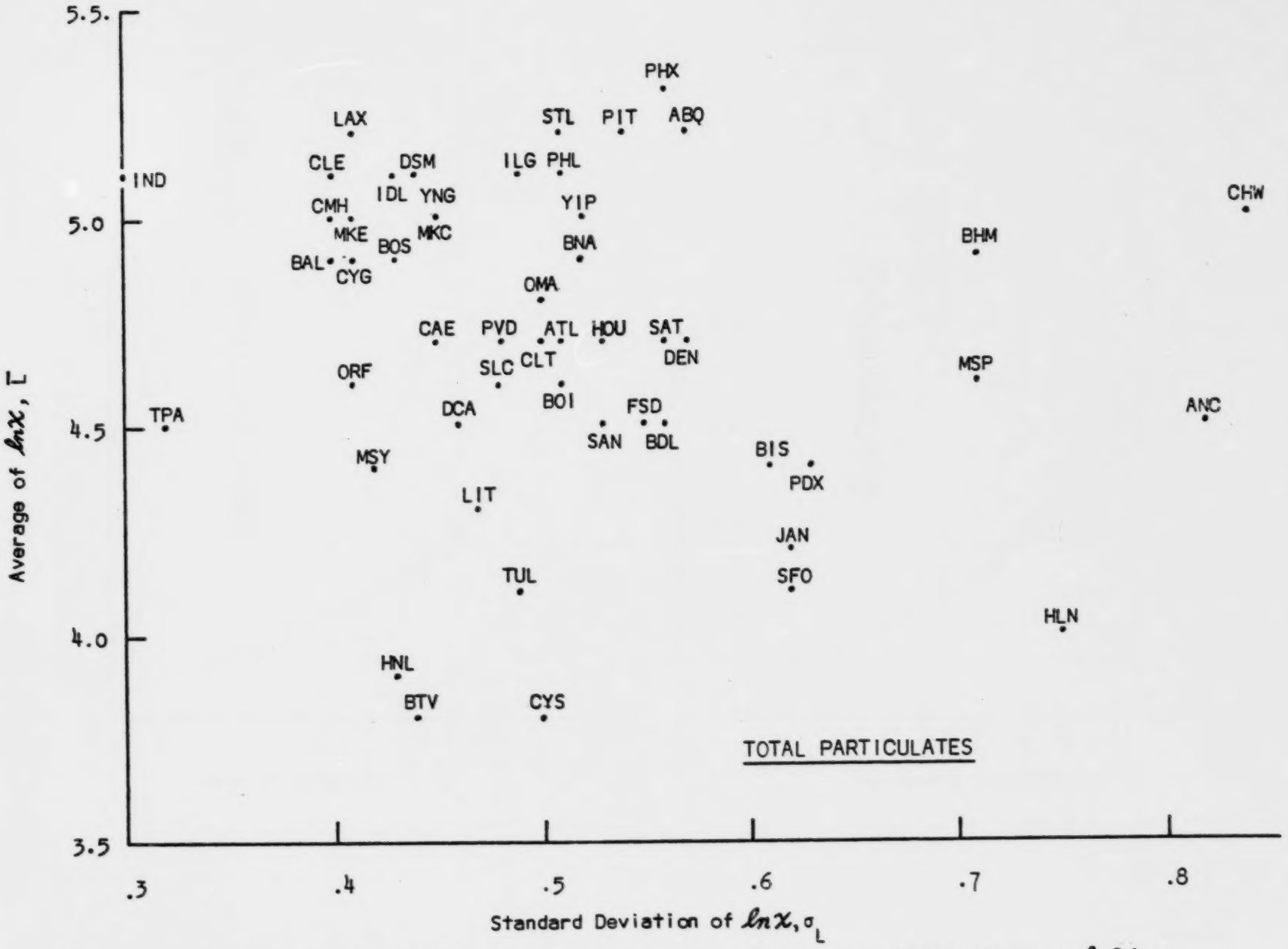



Fig. 3. Joint values of central tendency and dispersion of frequency distributions of $\ln x$ as measured by L and $\sigma_L$ for 51 urban stations. Data for period 1957 - 1959. (See Table 1 for station identification, e.g., DCA = Washington, D.C.)

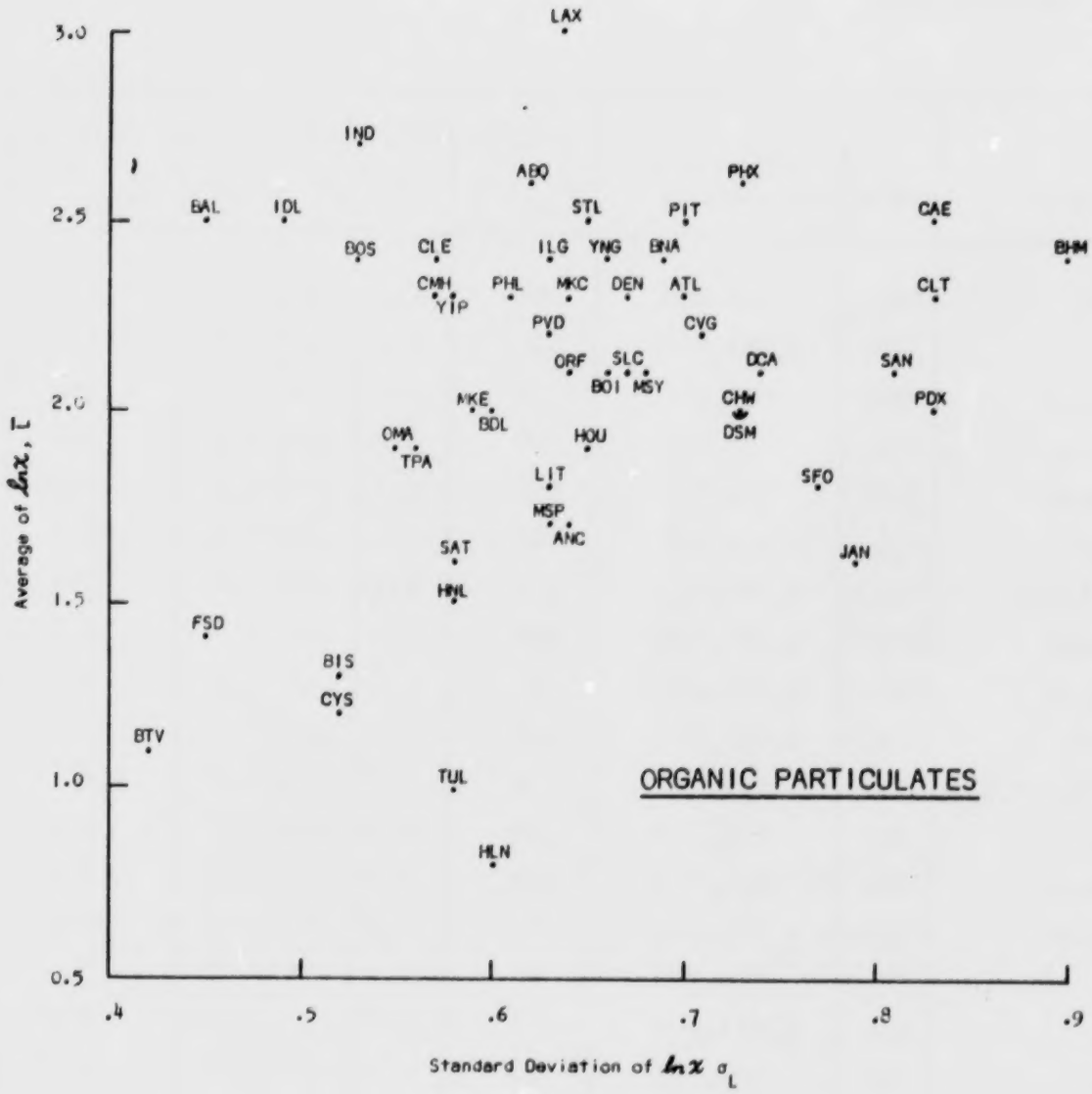



Fig. 4. Joint values of central tendency and dispersion of frequency distribution of $\ln x$, as measured by $\bar{L}$ and $\sigma_L$, for 51 urban stations. Data for period 1957 - 1959. (See Table 1 for station identification, e.g., DCA = Washington, D.C.)

TABLE 1

Values of median concentration ($\chi_{med}$), average of $\ln\chi$ ($\bar{L}$), standard deviation of $\ln\chi$ ($\sigma_L$), and average of $\ln\chi$ normalized to a standard population of $10^6$ ($\bar{L}^*$) for NASN measurements from 51 stations during period 1957 through 1959.

| Station | Code | Metropolitan Population[1] | Total Particulates | | | | Organic Particulates | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\chi_{med}$[2] | $\bar{L}$ | $\sigma_L$ | $\bar{L}^*$ | $\chi_{med}$[2] | $\bar{L}$ | $\sigma_L$ | $\bar{L}^*$ |
| Hartford | BDL | 763,600 | 92 | 4.5 | .56 | 4.5 | 7.4 | 2.0 | .60 | 2.1 |
| Boston | BOS | 2,913,500 | 134 | 4.9 | .43 | 4.7 | 11.1 | 2.4 | .53 | 2.2 |
| Providence | PVD | 804,300 | 100 | 4.7 | .48 | 4.7 | 8.4 | 2.2 | .63 | 2.2 |
| Burlington | BTV | 35,500 | 46 | 3.8 | .44 | 4.4 | 2.9 | 1.1 | .52 | 1.8 |
| Wilmington | ILG | 318,700 | 166 | 5.1 | .49 | 5.3 | 10.6 | 2.4 | .63 | 2.6 |
| New York | IDL | 15,404,300 | 166 | 5.1 | .43 | 4.6 | 12.3 | 2.5 | .49 | 3.0 |
| Philadelphia | PHL | 3,969,500 | 176 | 5.1 | .51 | 4.9 | 9.2 | 2.3 | .61 | 2.0 |
| Pittsburgh | PIT | 1,957,700 | 163 | 5.2 | .54 | 5.1 | 11.8 | 2.5 | .70 | 2.4 |
| Washington | DCA | 2,053,600 | 92 | 4.5 | .46 | 4.4 | 8.4 | 2.1 | .74 | 2.0 |
| Baltimore | BAL | 1,636,500 | 128 | 4.9 | .40 | 4.8 | 11.7 | 2.5 | .45 | 2.4 |
| Charlotte, N.C. | CLT | 272,700 | 108 | 4.7 | .50 | 4.9 | 11.1 | 2.3 | .83 | 2.6 |
| Norfolk | ORF | 574,900 | 108 | 4.6 | .41 | 4.7 | 7.8 | 2.1 | .64 | 2.2 |
| Charleston | CHW | 213,900 | 150 | 5.0 | .84 | 5.3 | 8.1 | 2.0 | .73 | 2.3 |
| Birmingham | BHM | 624,000 | 144 | 4.9 | .71 | 5.0 | 10.0 | 2.4 | .90 | 2.5 |
| Tampa | TPA | 356,200 | 90 | 4.5 | .32 | 4.7 | 7.0 | 1.9 | .56 | 2.1 |
| Atlanta | ATL | 1,011,100 | 112 | 4.7 | .51 | 4.7 | 8.8 | 2.3 | .70 | 2.3 |
| Jackson | JAN | 191,200 | 71 | 4.2 | .62 | 4.5 | 5.2 | 1.6 | .79 | 1.9 |
| Columbia | CAE | 213,400 | 104 | 4.7 | .45 | 5.0 | 12.5 | 2.5 | .83 | 2.8 |
| Nashville | BNA | 411,500 | 126 | 4.9 | .52 | 5.1 | 8.7 | 2.4 | .79 | 2.6 |
| Indianapolis | IND | 806,900 | 165 | 5.1 | .30 | 5.1 | 14.2 | 2.7 | .53 | 2.7 |
| Detroit | YIP | 3,838,500 | 135 | 5.0 | .52 | 4.8 | 9.7 | 2.3 | .58 | 2.0 |
| Cincinnati | CVG | 1,203,300 | 127 | 4.9 | .41 | 4.9 | 9.2 | 2.3 | .71 | 2.3 |
| Youngstown | YNG | 467,600 | 145 | 5.0 | .45 | 4.9 | 11.7 | 2.4 | .66 | 2.2 |
| Cleveland | CLE | 2,090,800 | 168 | 5.1 | .40 | 5.0 | 11.1 | 2.4 | .57 | 2.3 |

TABLE 1 - CONT.

| Station | Code | Metropolitan Population(1) | Total Particulates | | | | Organic Particulates | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\chi_{med}$ (2) | $\bar{L}$ | $\sigma_L$ | $\bar{L}$* | $\chi_{med}$ (2) | $\bar{L}$ | $\sigma_L$ | $\bar{L}$* |
| Columbus | CMH | 715,400 | 148 | 5.0 | .40 | 5.1 | 10.1 | 2.3 | .57 | 2.4 |
| Milwaukee | MKE | 1,240,700 | 133 | 5.0 | .41 | 5.0 | 7.6 | 2.0 | .59 | 2.0 |
| Des Moines | DSM | 261,900 | 164 | 5.1 | .44 | 5.3 | 7.5 | 2.0 | .73 | 2.3 |
| Minneapolis | MSP | 1,441,700 | 102 | 4.6 | .71 | 4.5 | 6.0 | 1.7 | .63 | 1.6 |
| Kansas City | MKC | 1,025,900 | 150 | 5.0 | .45 | 5.0 | 9.7 | 2.3 | .64 | 2.3 |
| St. Louis | STL | 2,050,800 | 182 | 5.2 | .51 | 5.1 | 11.8 | 2.5 | .65 | 2.4 |
| Omaha | OMA | 434,800 | 114 | 4.8 | .50 | 4.9 | 6.8 | 1.9 | .55 | 2.1 |
| Bismarck | BIS | 27,700 | 80 | 4.4 | .61 | 5.0 | 3.6 | 1.3 | .52 | 2.0 |
| Sioux Falls | FSD | 65,466 | 90 | 4.5 | .55 | 5.0 | 4.5 | 1.4 | .45 | 1.9 |
| Little Rock | LIT | 238,500 | 72 | 4.3 | .47 | 4.5 | 5.8 | 1.8 | .63 | 2.1 |
| New Orleans | MSY | 885,200 | 83 | 4.4 | .42 | 4.4 | 9.1 | 2.1 | .68 | 2.1 |
| Albuquerque | ABQ | 266,300 | 184 | 5.2 | .57 | 5.4 | 13.2 | 2.6 | .62 | 2.9 |
| Tulsa | TUL | 360,900 | 72 | 4.1 | .49 | 4.3 | 3.0 | 1.0 | .58 | 1.2 |
| Houston | HOU | 1,251,700 | 112 | 4.7 | .53 | 4.7 | 8.0 | 1.9 | .65 | 1.9 |
| San Antonio | SAT | 689,700 | 120 | 4.7 | .56 | 4.8 | 5.3 | 1.6 | .58 | 1.7 |
| Denver | DEN | 858,300 | 110 | 4.7 | .57 | 4.7 | 9.2 | 2.3 | .67 | 2.3 |
| Boise | BOI | 34,481 | 108 | 4.6 | .51 | 5.2 | 8.6 | 2.1 | .66 | 2.8 |
| Helena | HLN | 20,227 | 60 | 4.0 | .75 | 4.7 | 2.2 | 0.8 | .60 | 1.6 |
| Salt Lake City | SLC | 410,200 | 97 | 4.6 | .48 | 4.8 | 7.8 | 2.1 | .67 | 2.3 |
| Cheyenne | CYS | 43,500 | 52 | 3.8 | .50 | 4.3 | 3.6 | 1.2 | .52 | 1.8 |
| Anchorage | ANC | 80,000 | 93 | 4.5 | .82 | 4.9 | 5.4 | 1.7 | .64 | 2.2 |
| Phoenix | PHX | 619,600 | 215 | 5.3 | .56 | 5.4 | 11.3 | 2.6 | .73 | 2.7 |
| Los Angelex | LAX | 6,565,000 | 185 | 5.2 | .41 | 4.9 | 21.2 | 3.0 | .67 | 2.6 |
| San Francisco | SFO | 3,275,000 | 61 | 4.1 | .62 | 3.9 | 5.5 | 1.8 | .77 | 1.6 |
| San Diego | SAN | 890,000 | 88 | 4.5 | .53 | 4.5 | 7.7 | 2.1 | .81 | 2.1 |
| Honolulu | HNL | 486,400 | 53 | 3.9 | .43 | 4.0 | 5.2 | 1.5 | .58 | 1.6 |
| Portland | PDX | 731,200 | 83 | 4.4 | .63 | 4.5 | 7.2 | 2.0 | .83 | 2.1 |

(1) - 1960 Census

(2) - $\mu g/m^3$

The level and variability of air pollution in an isolated urban complex depend upon four major factors:

1. The source types (kinds of material emitted), their distribution in space, and the source strength.
2. Periodicities of source strength.
3. Local terrain as it influences channelling of atmospheric flow and turbulence generation.
4. Meteorological variables affecting transport, diffusion and deposition of airborne materials.

In addition, local variations of concentrations will occur because of non-uniformity of source behavior and distribution, terrain, and meteorological factors within the city. The location of the sampling point therefore also becomes a primary consideration in analyzing pollution data.

The knowledge and information necessary to relate the observed pollution levels at an urban station to the source, terrain, and meteorological variables are not now available, and a clear delineation of the reasons for observed differences of pollution experience among cities is not presently possible. However, two interesting partial explanations for these observed differences have arisen from these studies and are presented as areas for further explorration. The first of these relates the observed median value of concentration to the metropolitan population. The second delineates geographical areas of anomalous pollution levels which coincide approximately with major terrain features, such as plains and mountains.

A. The Dependence of $\bar{L}$ and $\sigma_L$ upon Metropolitan Populations.

Recent studies (e.g., Hilst[2]) have shown that the cumulative

---

2. Hilst, G. R., 1961: Transactions of Symposium, "Air over Cities," USPHS, R. A. Taft Sanitary Engineering Center. (In Press.)

concentration of pollutants in air passing over an <u>area</u> source increases as a power of the distance so traversed. If the population is proportional to the square of a characteristic length of the city, and the pollutant source strength per inhabitant is a constant, then the concentration of pollutant is proportional to a power of the total population.

$$\chi \propto P_o^{\,B} \qquad (1)$$

Taking the median value of concentration $\chi_{med}$ as a representative measure of the pollution experience,

$$\bar{L} \quad \ln \chi_{med} \quad \propto \quad B \ln P_o \qquad (2)$$

Then

$$\bar{L} = \bar{L}^* + B \ln Po/Pn \qquad (3)$$

Where

$\bar{L}$ = observed average ln of concentration

$\bar{L}^*$ = normalized $\bar{L}$

$P_o$ = observed metropolitan population

$P_n$ = base population for normalization

Equation (3) provides the basis for normalizing pollution levels to a base reference population.

Implicit in the development of this basis for normalizing pollution levels are two further assumptions, 1) that sources are uniformly distributed, and 2) the effects of pollution abatement control in reducing source strengths are uniform among the cities compared. Neither of these assumptions can be substantiated; they are, in fact, denied by evidence to be presented later.

Despite these shortcomings it appears to be worthwhile to attempt an approximate normalization of $\bar{L}$ on the basis of population.

Observed values of $\bar{L}$ are plotted as a function of metropolitan population in Figures 5 and 6. Values of $\bar{L}$ for both total and organic particulates show a definite partial dependence upon metropolitan population. This dependence is quantized by fitting these data to the regression equation suggested by equation (3). When this is done we find that 35% of the variance of $\bar{L}$ for both total particulates and organic particulates is accounted for by $\ln P_o$. The value of B is .169 for total particulates and .199 for organic particulates. (In fitting the regression equation for $\bar{L}$ vs $\ln P_o$ for total particulates, San Francisco (SFO) was deleted because of its obviously anomalous behavior in relation to other cities. Honolulu appears to be somewhat anomalous in $\bar{L}$ vs $\ln P_o$, also, but was retained.)

$\sigma_L$ is invariant for this transformation of $\bar{L}$, hence no normalization of the <u>dispersion</u> of the frequency distribution on the basis of population is called for, although, of course, the range of $\chi$ embraced by $\bar{L} \pm \sigma_L$ increases or decreases as $\bar{L}$ increases or decreases.

This significant dependence of $\bar{L}$ on the metropolitan population has been used to normalize the median concentration, or $\bar{L}$, to a unit population. A metropolitan population of 1,000,000 people has been chosen as this unit base and values of $\bar{L}$ have been normalized to $P_n = 10^6$ by

$$\bar{L}^* = \bar{L} - B \ln P_o/10^6, \qquad (4)$$

where B = .169 for total particulates and .199 for organic particulates. These values of $\bar{L}^*$ are tabulated in Table 1. Some 65% of the variability of $\bar{L}$ among the cities remains, of course, and must be attributed to an unknown dependence upon source variability, terrain variability, meteorological variability, and sampling location.

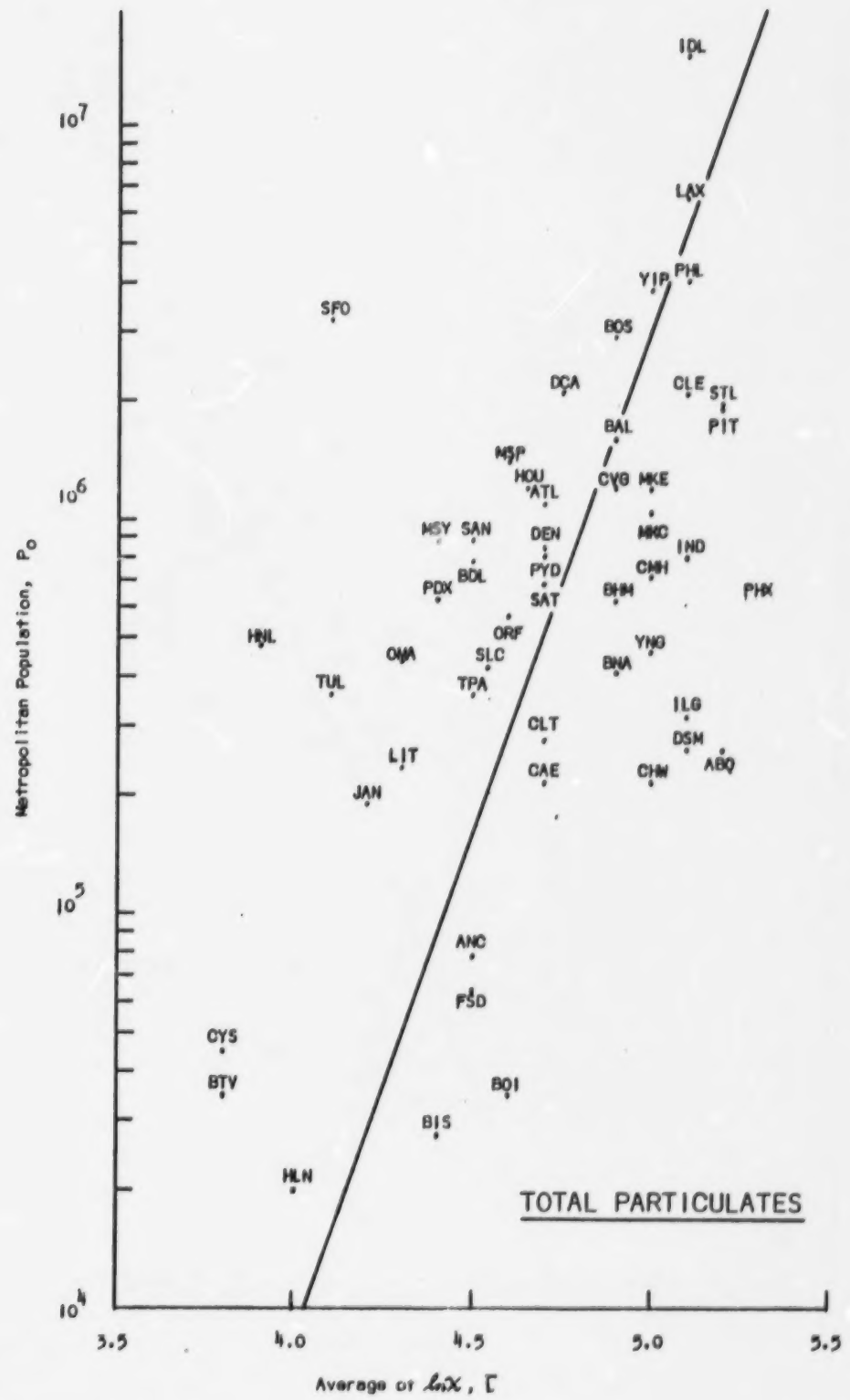



Fig. 5. Average of $\ln X$ as function of metropolitan population. Regression line given by $\bar{L} = 4.81 + .17 \ln P_o/10^6$.

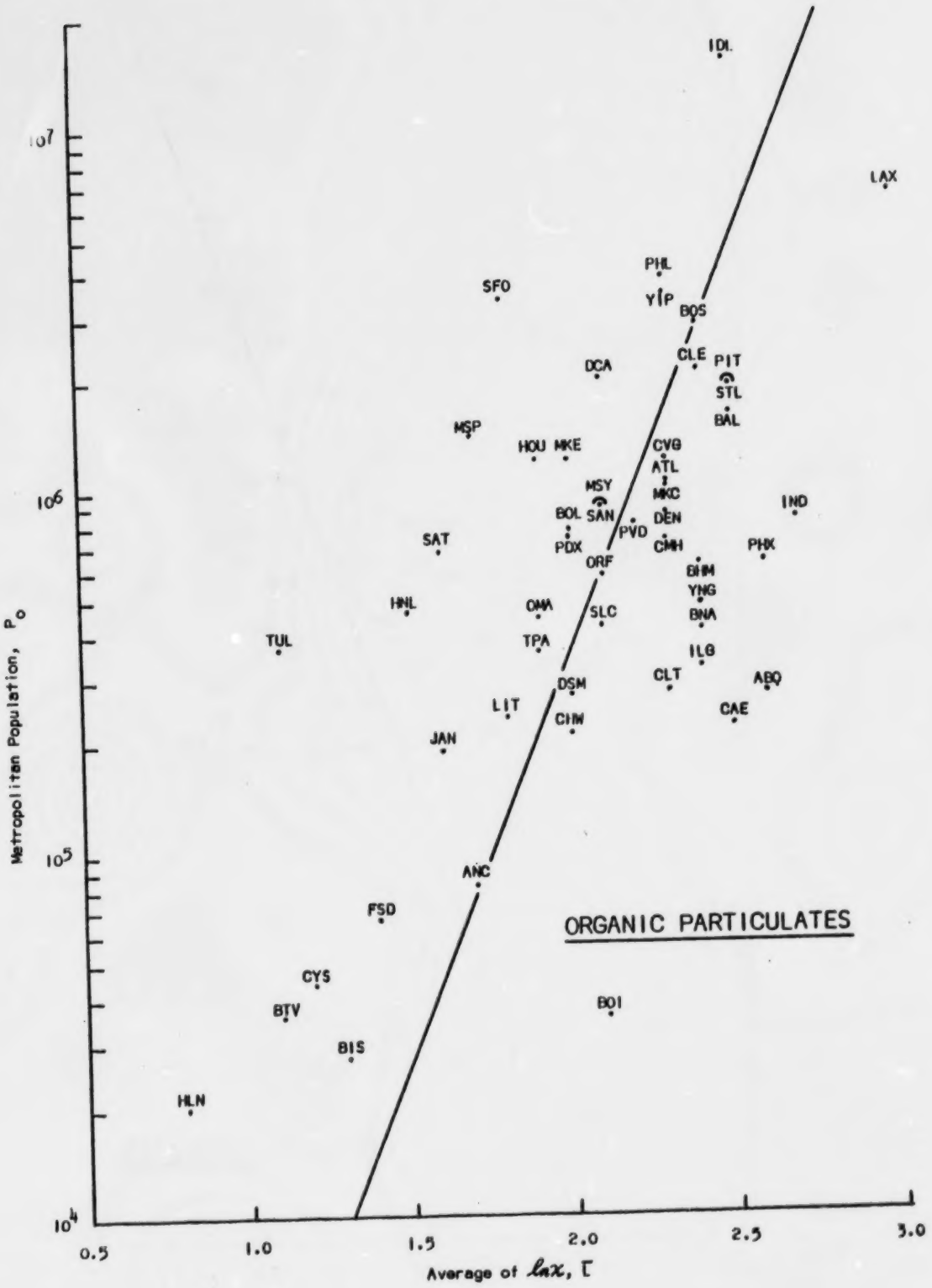



Fig. 6. Average of $\ln x$ as function of metropolitan population. Regression line given by $\bar{L} = 2.19 + .20 \ln P_o/10^6$.

B. The Dependence of $\bar{L}$ and $\sigma_L$ upon Geographical Location.

The observed values of $\bar{L}, \sigma_L$ and $\bar{L}^*$ may now be used to examine the geographic variations of these measures of pollution experience. A direct examination of these variations may be obtained by plotting the values of $\bar{L}$, $\sigma_L$ and $\bar{L}^*$ on base maps and delineating areas of similarity. In doing so it is essential that the reader remember these values are appropriate to urban areas and have meaning with regard to non-urban areas only insofar as they indicate the expected levels of pollution should a city, of the types included in this study, be built in the presently rural areas.

The geographical variability of $\bar{L}, \sigma_L$ and $\bar{L}^*$ are shown in Figures 7 - 9 for total particulates and Figures 10 - 12 for organic particulates.

Examination of the observed values of $\bar{L}$ in both Figures 7 and 10 show striking similarities in the distribution of relative values of organic and inorganic pollution concentration across the United States. Four rather distinct areas stand out; 1) An area of relatively low values of $\bar{L}$ over the Pacific Northwest, including northern California, eastward across the northern Rocky Mountain states and over the high plains area, the northern Great Lakes and extending into northern New England; 2) An area of relatively high concentrations over the Southwest U. S., 3) An area of relatively low concentrations over much of the Atlantic Seaboard, the Gulf Coast and extending northwestward over Louisiana, Arkansas, Eastern Texas and much of Oklahoma and Kansas; 4) an area of relatively high concentrations over the industrialized areas of the Midwest, East, and South.

A similar comparison of $\sigma_L$ for these two classes of pollution (Figures 8 and 11) fails to show such similarity of patterns, although some remnants remain, particularly in the Northwest and Southeast. The variability of $\sigma_L$ for total particulate concentrations (Figure 8) is quite small over most of the continental United States. Only a relatively high degree of variability

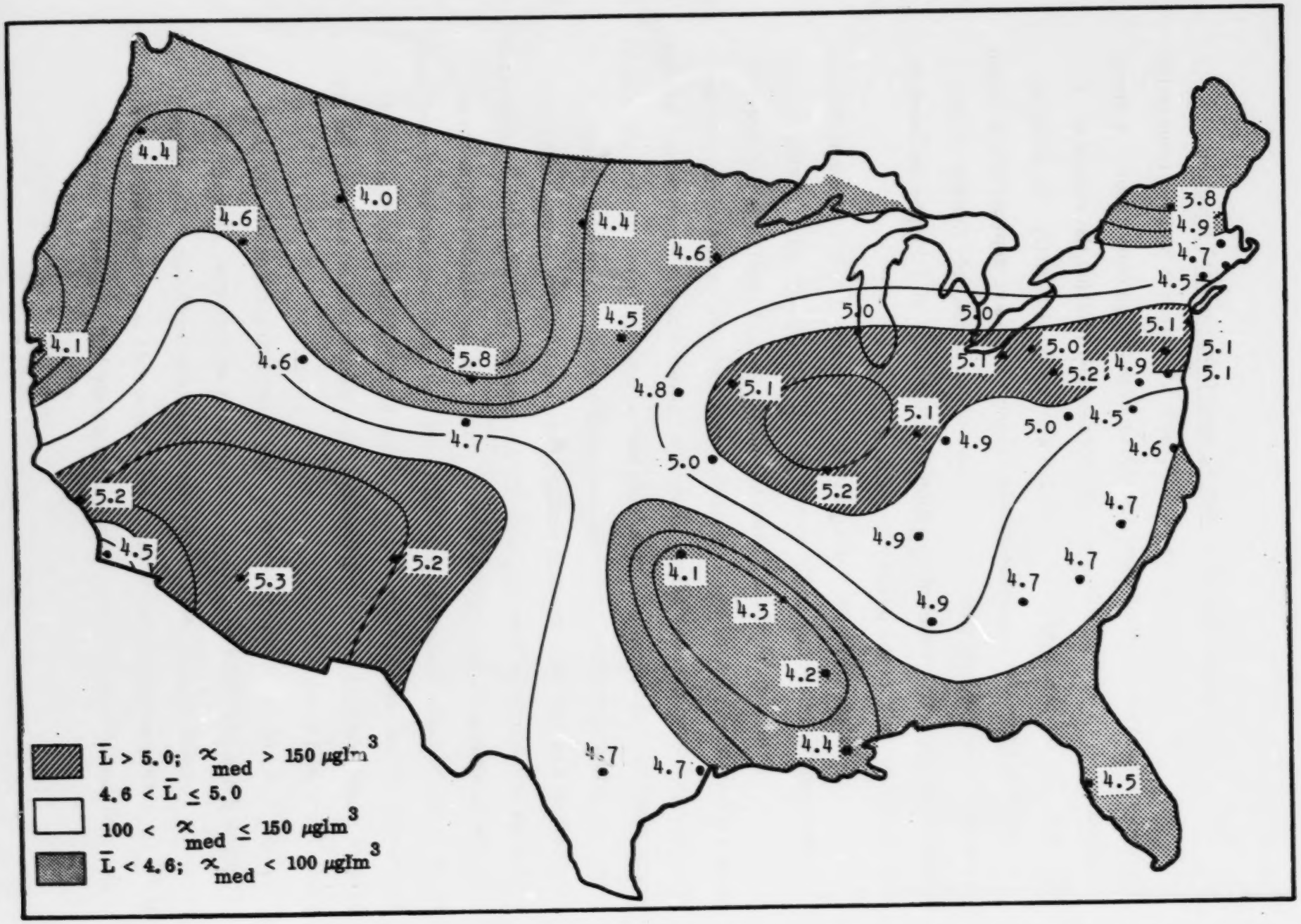



**Fig. 7.** Values of average of $\ln$ of total particulate concentrations ($\bar{L}$) observed in the Continental United States during 1957 through 1959. (National Air Sampling Network data)

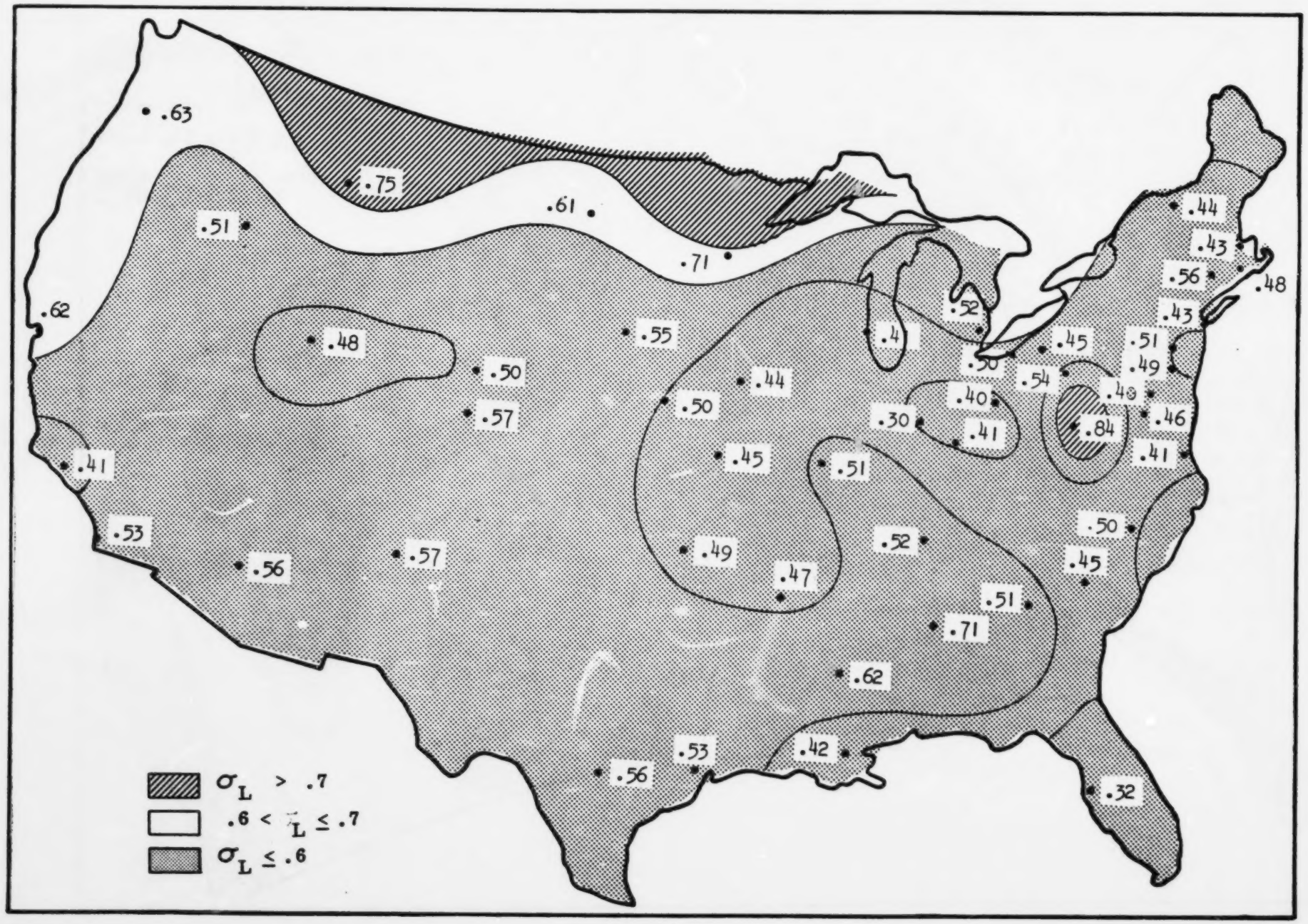



Fig. 8. Values of the standard deviation of $\ell n$ of total particulate concentrations ($\sigma_L$) observed in Continental United States during 1957 - 1959. (National Air Sampling Network Data)

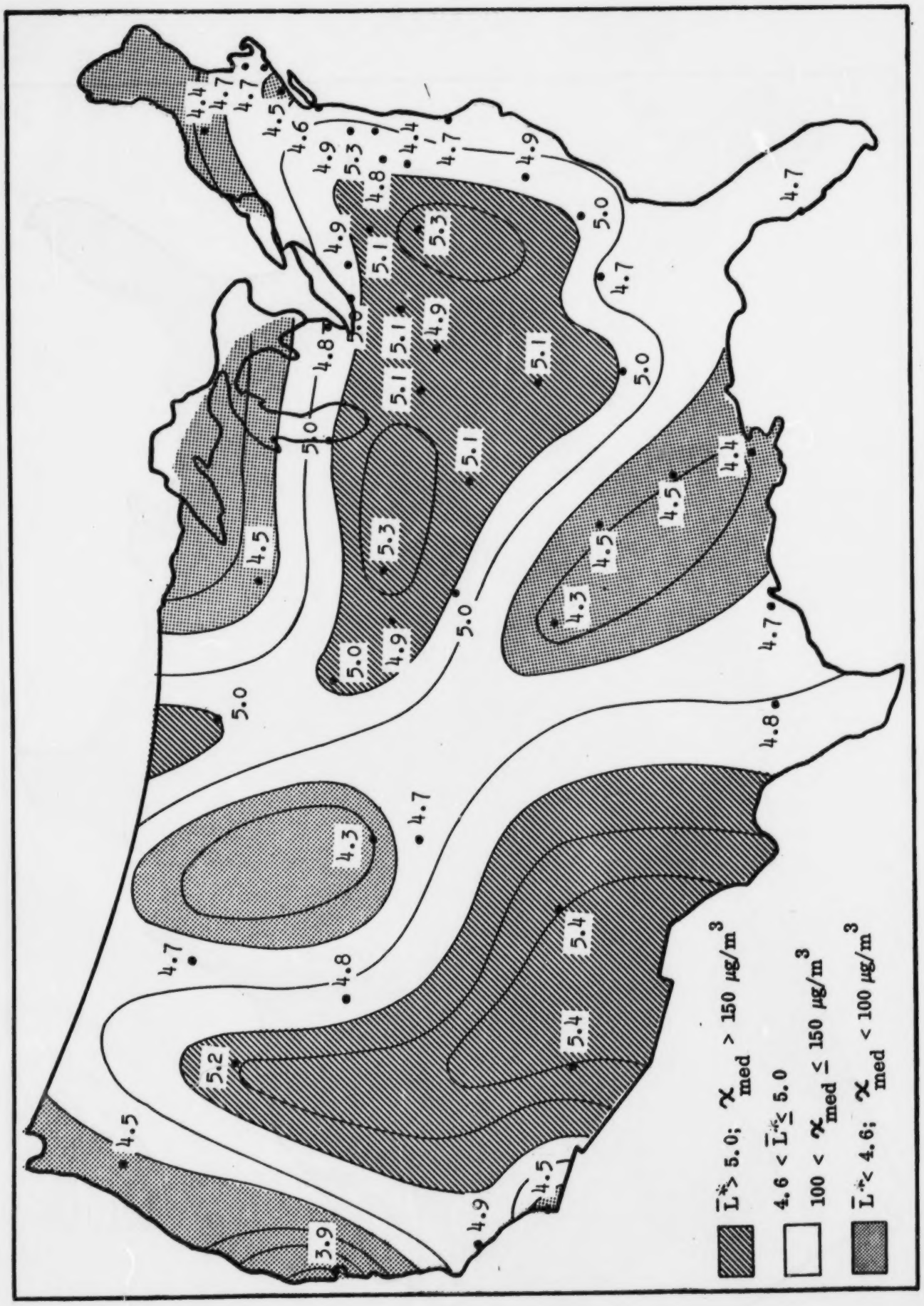



Fig. 9. Values of average of $\mathcal{A}_n$ total particulate concentrations ($\bar{L}^*$) normalized to a base metropolitan population of $10^6$, as observed during period 1957 - 1959. (National Air Sampling Network Data.)

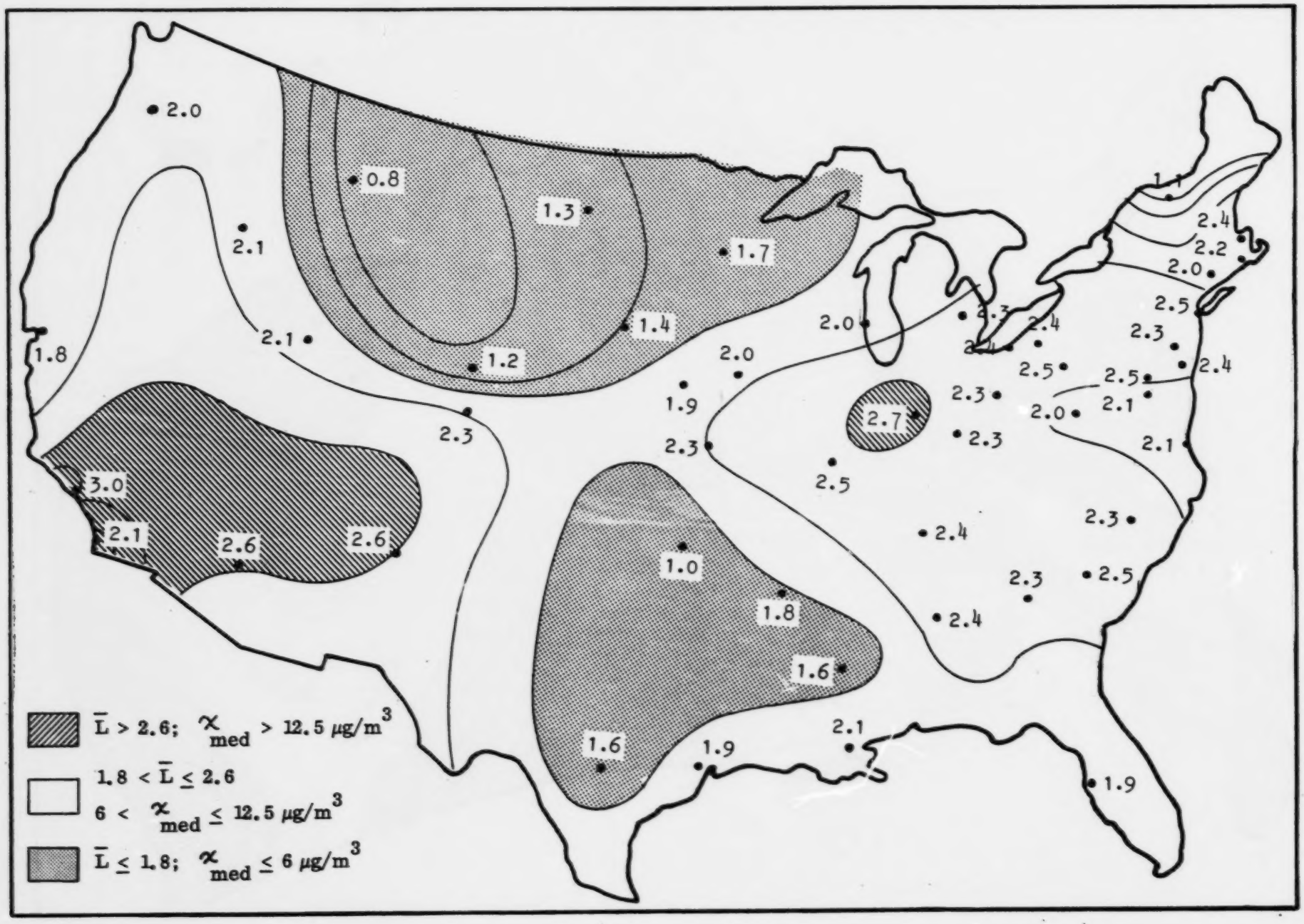



Fig. 10. Values of average of $\ln$ of organic particulate concentrations $(\bar{L})$ observed in Continental United States during 1957-1959. (National Air Sampling Network Data.)

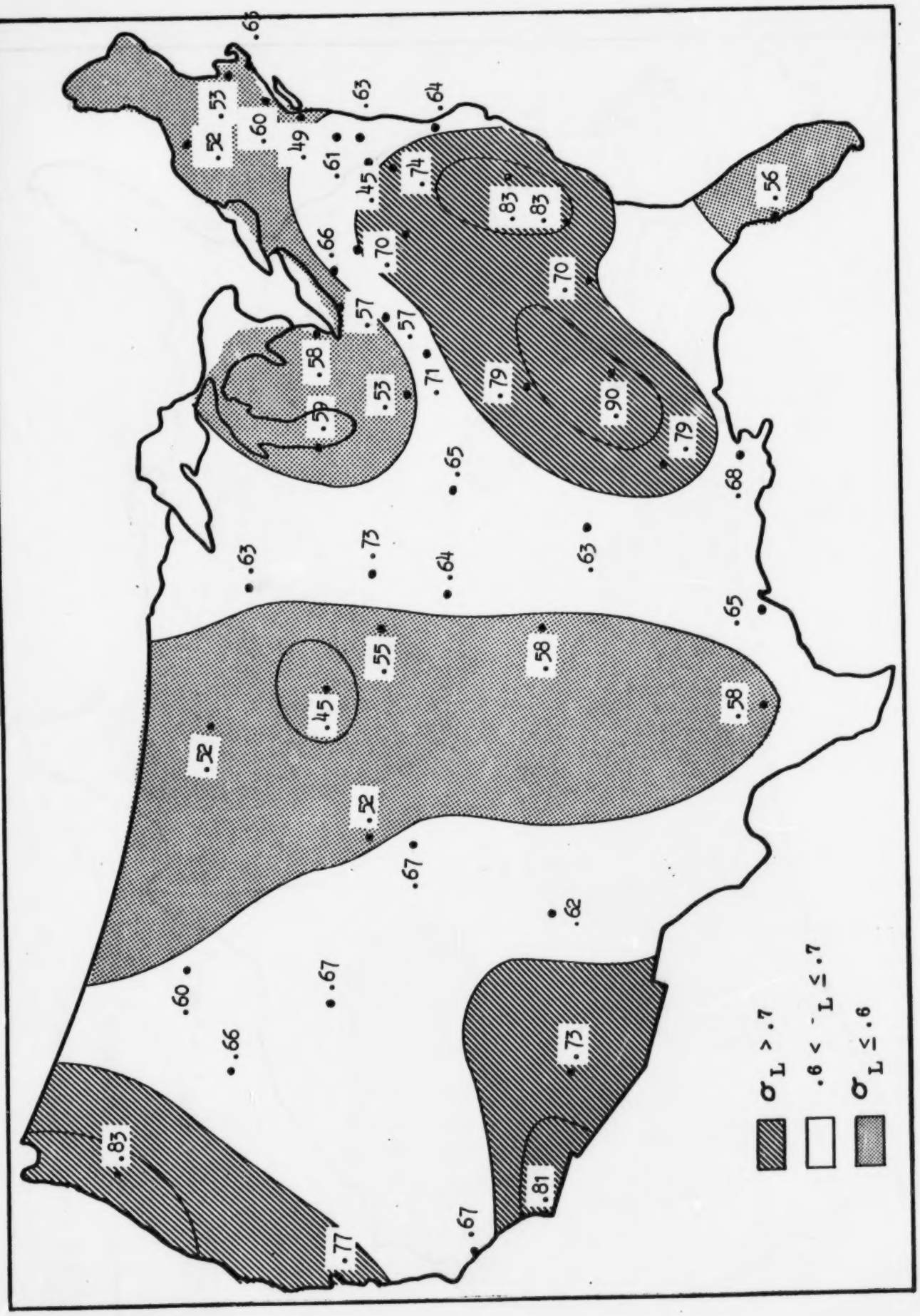



Fig. 11. Values of standard deviation of $\sigma_L$ of organic particulate concentrations observed in Continental United States during 1957 - 1959. (National Air Sampling Network Data.)

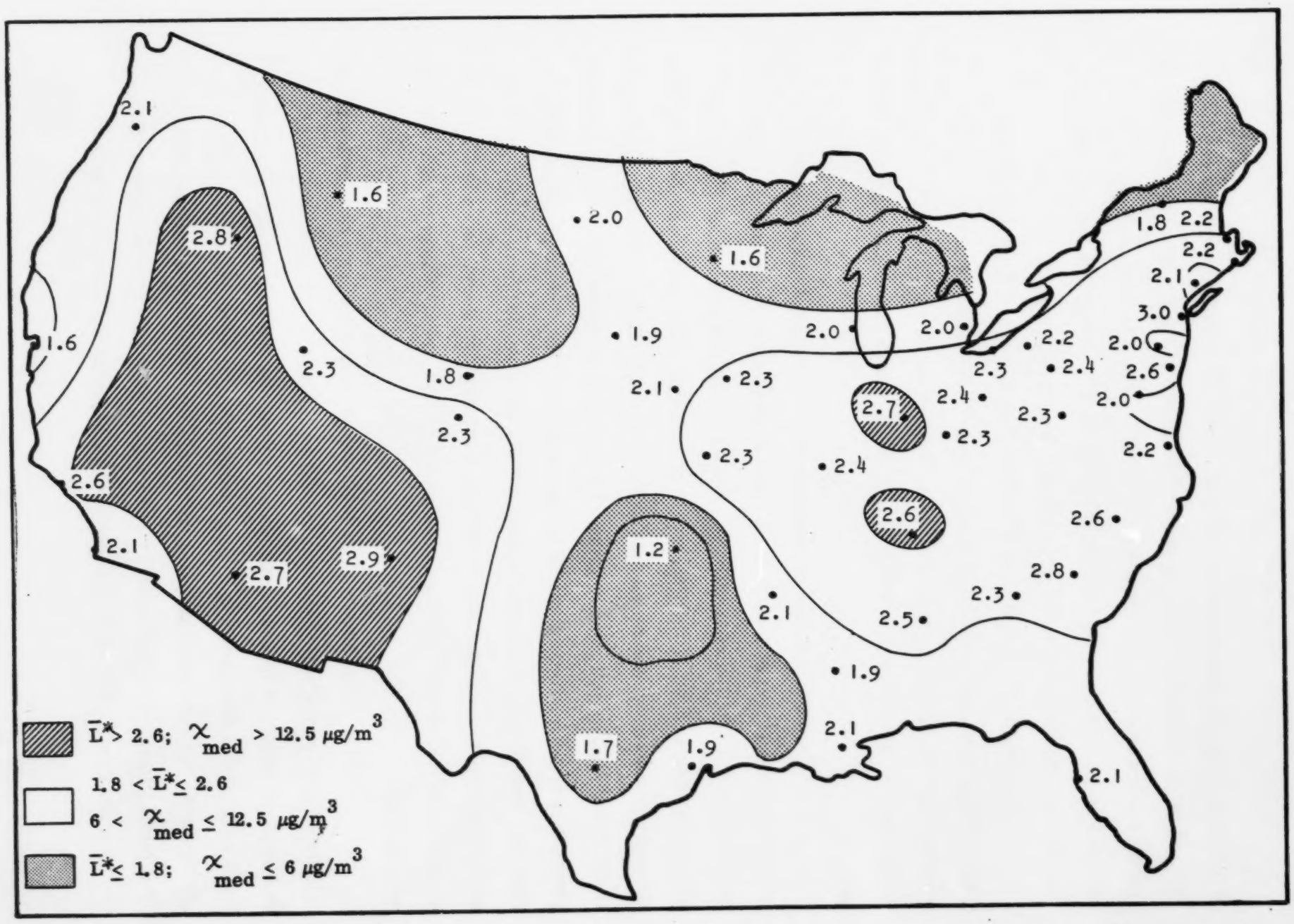



Fig. 12. Values of average of $ln$ of organic particulate concentrations normalized to a base population of $10^6$ as observed during period 1957 - 1959. (National Air Sampling Network Data.)

extending along the northern tier of states from Washington to Minnesota stands out. The variations of $\sigma_L$ for organic particulates (Figure 11) are considerably greater and more regionalized with relatively small values over much of the Great Plains, the Great Lakes and New England, and larger values over the Northwest, Southwest, and the Southeast, excluding most of Florida.

These differences of $\sigma_L$ between total particulates and organic particulates reemphasizes the earlier suggestion that the sources for the organic and inorganic particulates are different since the levels of the pollutants measured by the Network are the result of the same meteorological and terrain effects.

The distributions of median values of concentration normalized to a base metropolitan population of $10^6$ are shown in Figure 9 and 12. Comparison with Figures 7 and 10, respectively, show that no major changes in the geographical distribution of median levels of concentration are occasioned by this normalization except for an accentuation of the potential for high concentrations in the inter-Rocky Mountain region. The number of cities and the sample size preclude any definitive statement in this regard. Yet, the implication is there that the inter-Rocky Mountain region is fully as susceptible to urban pollution as the presently industrialized Midwest, East and South and, because the degree of industrialization per capita is now small in this area, it may be much more susceptible.

While it is tempting, at least for a meteorologist, to identify areas of relatively high and low pollution values with terrain and meteorological features, it would be foolhardy to advance these as the only or primary basis for the observed distributions. Most assuredly the meteorological factors are at work; but one can recognize that variations of sources, as related to total population, industrial activity per capita, automobiles per capita, etc., are also involved. The differentiation and integration of these multiple and

complex factors must await much more fundamental study and measurement. Only then will we be able to answer conclusively, "What pollution levels can we expect if we enlarge this city, or build a new city of a given size and complexion of population and industry, in a particular locality?" In the meantime, we have some guidelines.

## III. THE PREDICTABILITY OF PARTICULATE POLLUTION LEVELS FROM CONVENTIONAL METEOROLOGICAL MEASUREMENTS

The primary question asked of the data in the analyses reported in Volume I may be phrased as follows:

> "Given the meteorological measurements listed in Appendix A, how well can one predict whether the associated particulate pollution will be greater than or less than the median value appropriate to the sampling site used in the National Air Sampling Network?"

This question must be asked individually for each station included in the Network until such time as necessary and sufficient criteria for a universal prediction system have been developed.

The second question, closely allied with the first, is:

> "To what extent do the individual meteorological measurements assist in the discrimination between (or prediction of) pollution levels greater than or less than the median value?"

This question drives directly to the basic question, "How is the pollution level related to atmospheric motions, processes, and physical state parameters?" However, the answers sought in the present analyses are subject to two severe restrictions, 1) only qualitative discrimination of two groups of pollution level is looked for, and 2) the answers are sought within the context of incomplete and probably inappropriate measurements. The first of these restrictions is necessary because of the limited data sample; analysis techniques which can discriminate among multi-groups of pollution level are available but require much more extensive data collections.

The second restriction stems primarily from a lack of comprehensive quantitative knowledge of the multiple physical processes involved in air pollution; knowledge from which the measurements required to predict the pollution level could be specified. It is, of course, precisely this latter knowledge which is

sought through basic research programs. The present analyses may be viewed as a contribution to such a research effort in the sense that information inherent in those measurements which have been made can be discerned and the residual variability can be specified. Further, the present analyses provide valuable first approximations to the prediction of anomalous pollution levels, given the meteorological information employed here.

A. Predictability

As explained more fully in Volume I, the two-group discriminant analysis technique employed in these studies compares in a systematic manner the variability of the seasonal and meteorological measurements to the occurrence of pollution levels (HIGH) and below (LOW) the median value of the period of record. This is done on a station-by-station basis and the output is a ranking of the seasonal and meteorological parameters in the order of decreasing contribution to the discrimination between these two pollution classes. In the present use of this technique the seasonal (day of the year) variabilities were analyzed first in such a way as to remove, as far as possible, their non-meteorological contribution to the occurrence of HIGH and LOW pollution levels. The seasonal meteorological variability was retained.

Following this removal of seasonal non-meteorological variability, the analysis technique was forced to evaluate the contribution of seasonal meteorological variability, as identified by the day of the year, and the contribution of the day-of-the-week (a non-meteorological variable) to the discrimination of HIGH and LOW pollution level occurrences. (Initially, the contribution of these two mixed variables was lumped under a single heading, seasonal base (SB). Later their contributions were separated.) The analysis then proceeded to select from among the eleven current meteorological variables those which demonstrated greatest further discrimination of HIGH and LOW pollution levels. The complete results of this analysis are listed in Table 24 of Volume I of this Report.

The ability of predictors to discriminate between HIGH and LOW pollution levels is measured in terms of the reduction of mean square error of the choice of pollution level by these predictors. When no discriminating information is available the mean square error (MSE) is 0.25, the maximum value this quantity can assume.

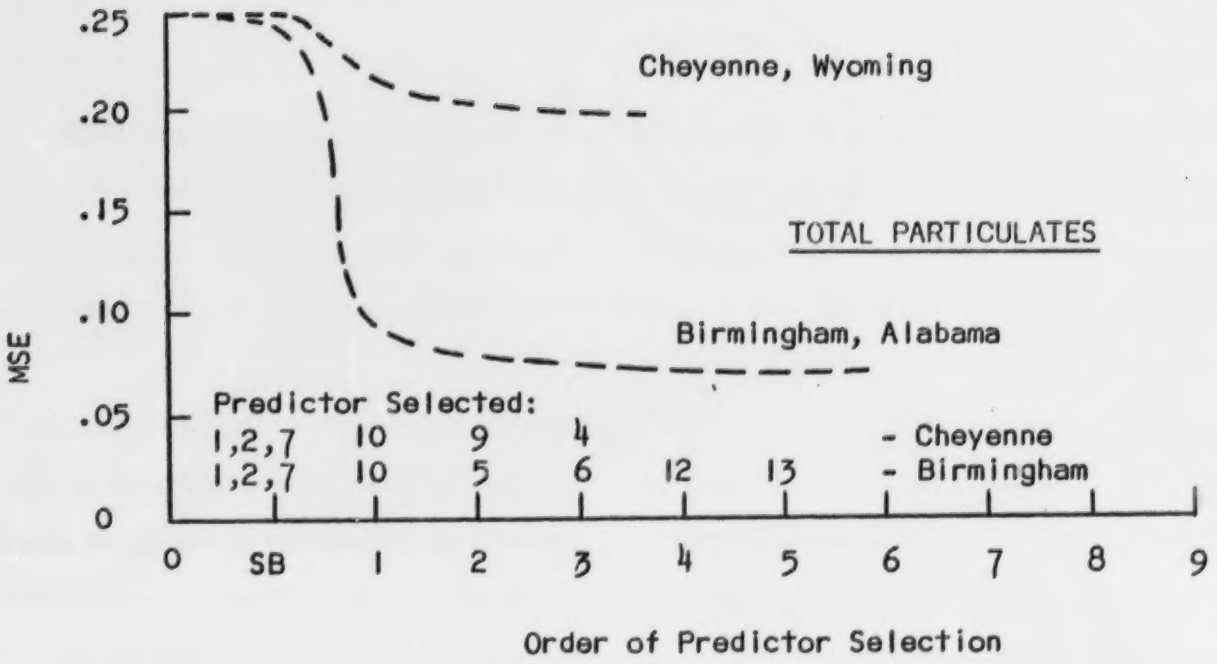
In order to illustrate this information graphically, the MSE for the stations of maximum and minimum predictability for both total particulates and organic particulates has been plotted in Figure 13 as a function of the predictors chosen sequentially by the two-group discriminant analysis. These curves show the reduction in MSE attributable to the seasonal base (SB) as defined above, and that which was due to each meteorological parameter chosen. The minimum mean square error is achieved after only a few (generally five to eight) of the eleven meteorological predictors have been selected; the information content of those predictors not selected is either redundant or non-existent.

In order to facilitate predictor designation, they have been identified by the same number code used in Volume I. The key to this code is listed in Table 2.

The predictors chosen and their relative contribution to the discrimination between HIGH and LOW pollution levels, examples of which are in Figure 13, are, of course, appropriate only to the individual stations. Similar analyses of the predictability and control of pollution levels by seasonal and meteorological parameters can be derived for any station of interest from the data presented in Table 24 of Volume I. This exercise is left to the reader, since interests will vary, and attention is turned here to comparison and interpretation of the results for all stations.

B. <u>The Contributions of Seasonal Meteorological Variability and Day-of-the-Week to Two-Group Pollution Prediction.</u>

As noted previously, the combined discriminatory information con-



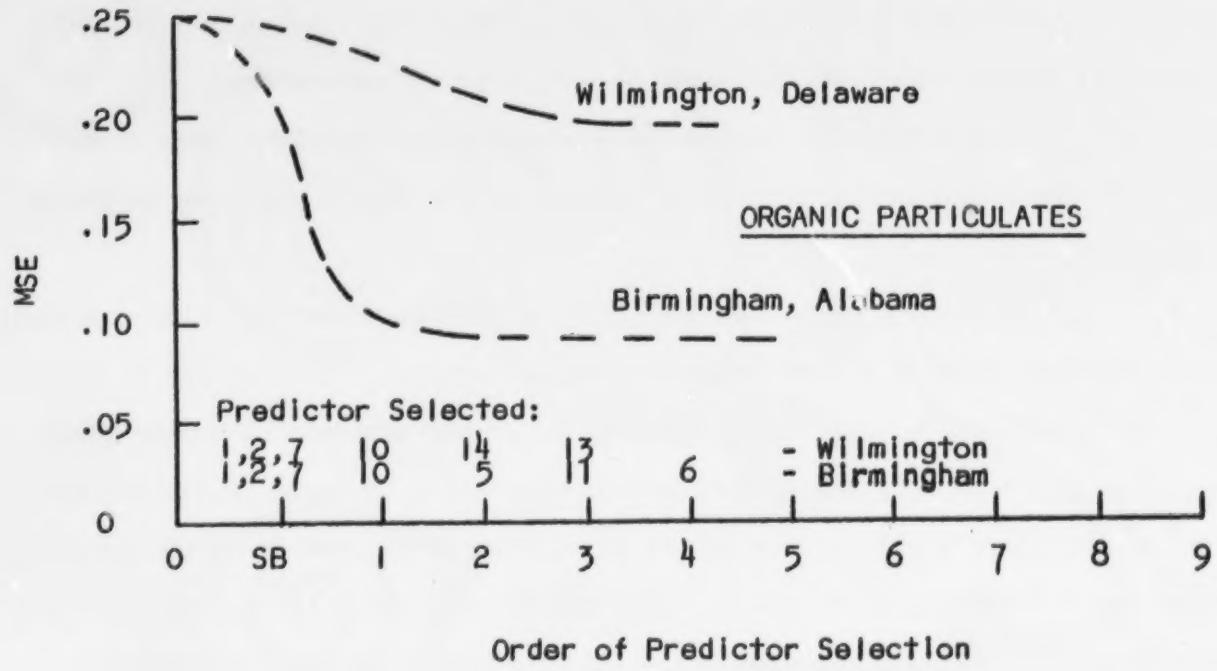




Fig. 13. **Reduction of mean square error (MSE) of discrimination between High and Low values of total and organic particulates for each predictor selected by two-group MDA. Predictors listed at bottom of figures were selected in order shown. (See Table 2 for identification of predictors).**

TABLE 2

Predictor Identification Code

| Code Number | Predictor |
|---|---|
| 1 | cosine wt (t=day of year) |
| 2 | sine wt |
| 3 | cube root of mean wind speed |
| 4 | temperature range |
| 5 | maximum temperature |
| 6 | minimum temperature |
| 7 | day of the week |
| 8 | amount of precipitation |
| 9 | hours of precipitation |
| 10 | wind direction |
| 11 | wind speed - 2 levels |
| 12 | relative humidity |
| 13 | cloudiness |
| 14 | variability of wind direction |

tained in the seasonal meteorological variability and in the day-of-the-week was originally lumped together, since primary interest centered in the information content of current meteorological variables. Later analyses pointed up the desirability of separating the contribution of these two factors since day-of-the-week was the only predictor which could be reasonably assumed independent of the current meteorological conditions and therefore a partial measure of souce variability.

The original combined contribution of these two variables, as taken

from Table 24 of Volume I, (SB), and their separate contributions, all in terms of the reduction of the mean square error of pollution level discrimination, are listed in Table 3. Inspection of this Table shows that in a vast majority of the cases the entire contribution of discrimination stems from the day-of-the-week. In order to highlight and facilitate this comparison, the relative contribution of these two variables is shown graphically in Figures 14 and 15. These Figures show that the seasonal meteorological contribution is more frequent and larger for organic particulates than for total particulates but, with a few notable exceptions, this contribution is small in both cases.

The contribution of the day-of-the-week will be discussed again in a later section on non-meteorological variables. However, for the purposes of the remainder of this Section, the combined contribution of seasonal weather and day-of-the-week may be used, with the realization that the bulk of this contribution lies in the day-of-the-week.

The geographic distribution of reduction in mean square error of pollution level selection accomplished by these variables ($\Delta$MSE(SB)) is shown in Figures 16 and 17. As might be expected from the fact that day-of-the-week contributes so prominently to $\Delta$MSE(SB), the regions of high contribution coincide with the more heavily populated and industrialized regions of the United States. The notable exception to this is the north Atlantic coastal region where heavy industrialization is not matched by a large value of $\Delta$MSE(SB). This may be due either to the lack of a uniform weekly cycle in the operation of these industries or to a relatively large contribution from non-industrial sources which operate with no fixed weekly pattern. Whatever the reason, and this must be determined by further measurements of source contributions and their variability, it is evident that over a very sizeable fraction of these urban areas the mean square error of choice between HIGH and LOW pollution levels may be reduced 20% simply by noting the day of the week.

TABLE 3

Background variations in pollution due to day of week and seasonal effects of weather[1].

| Station Name | Call Letters | Total Particulate | | | Organic Concentration | | |
|---|---|---|---|---|---|---|---|
| | | SB | D | W=SB-D | SB | D | W=SB-D |
| Hartford | BDL | .012 | .016 | .000 | .006 | .002 | .004 |
| Boston | BOS | .000 | .002 | .000 | .005 | .009 | .000 |
| Providence | PVD | .009 | .009 | .000 | .011 | .009 | .002 |
| Burlington | BTV | .014 | .019 | .000 | .020 | .025 | .000 |
| Wilmington | ILG | .000 | .006 | .000 | .005 | .004 | .001 |
| New York | IDL | .031 | .037 | .000 | .030 | .036 | .000 |
| Philadelphia | PHL | .027 | .016 | .011 | .062 | .043 | .019 |
| Pittsburgh | PIT | .026 | .023 | .003 | .024 | .009 | .015 |
| Washington | DCA | .017 | .014 | .003 | .015 | .018 | .000 |
| Baltimore | BAL | .032 | .028 | .004 | .028 | .028 | .000 |
| Charlotte | CLT | .024 | .025 | .000 | .023 | .003 | .020 |
| Norfolk | ORF | .024 | .027 | .000 | .023 | .023 | .000 |
| Charleston | CHW | .044 | .049 | .000 | .031 | .029 | .002 |
| Birmingham | BHM | .003 | .011 | .000 | .045 | .047 | .000 |
| Tampa | TPA | .026 | .028 | .000 | .018 | .004 | .014 |
| Atlanta | ATL | .057 | .042 | .015 | .053 | .045 | .008 |
| Jackson | JAN | .014 | .011 | .003 | .015 | .015 | .000 |
| Columbia | CAE | .000 | .007 | .000 | .003 | .009 | .000 |
| Nashville | BNA | .024 | .028 | .000 | .019 | .018 | .001 |
| Indianapolis | IND | .045 | .050 | .000 | .065 | .046 | .019 |
| Detroit | YIP | .038 | .041 | .000 | .031 | .034 | .000 |
| Cincinnati | CVG | .014 | .018 | .000 | .040 | .035 | .005 |
| Youngstown | YNG | .038 | .018 | .020 | .007 | .004 | .003 |
| Cleveland | CLE | .038 | .019 | .019 | .002 | .009 | .000 |
| Columbus | CMH | .025 | .021 | .004 | .035 | .028 | .007 |
| Milwaukee | MKE | .020 | .026 | .000 | .000 | .000 | .000 |

TABLE 3
cont.

| Station Name | Call Letters | Total Particulate | | | Organic Concentration | | |
|---|---|---|---|---|---|---|---|
| | | SB | D | W=SB-D | SB | D | W=SB-D |
| Des Moines | DSM | .044 | .030 | .014 | .035 | .035 | .000 |
| Minneapolis | MSP | .009 | .008 | .001 | .032 | .036 | .000 |
| Kansas City | MKC | .022 | .024 | .000 | .006 | .011 | .000 |
| St. Louis | STL | .048 | .048 | .000 | .063 | .059 | .004 |
| Omaha | OMA | .052 | .058 | .000 | .015 | .022 | .000 |
| Bismarck | BIS | .013 | .017 | .000 | .012 | .018 | .000 |
| Sioux Falls | FSD | .032 | .030 | .002 | .012 | .010 | .002 |
| Little Rock | LIT | .006 | .005 | .001 | .006 | .011 | .000 |
| New Orleans | MSY | .052 | .027 | .025 | .073 | .045 | .028 |
| Albuquerque | ABQ | .053 | .042 | .011 | .041 | .042 | .000 |
| Tulsa | TUL | .000 | .007 | .000 | .015 | .019 | .000 |
| Houston | HOU | .012 | .009 | .003 | .013 | .007 | .006 |
| San Antonio | SAT | .045 | .043 | .002 | .000 | .001 | .000 |
| Denver | DEN | .007 | .002 | .005 | .013 | .016 | .000 |
| Boise | BOI | .018 | .017 | .001 | .020 | .021 | .000 |
| Helena | HLN | .008 | .002 | .006 | .000 | .006 | .000 |
| Salt Lake City | SLC | .013 | .015 | .000 | .001 | .008 | .000 |
| Cheyenne | CYS | .000 | .004 | .000 | .034 | .012 | .022 |
| Anchorage | ANC | .010 | .003 | .007 | .009 | .009 | .000 |
| Phoenix | PHX | .020 | .022 | .000 | .023 | .018 | .005 |
| Los Angeles | LAX | .032 | .026 | .006 | .030 | .013 | .017 |
| San Francisco | SFO | .000 | .000 | .000 | .020 | .016 | .004 |
| San Diego | SAN | .050 | .033 | .017 | .044 | .015 | .029 |
| Honolulu | HNL | .021 | .028 | .000 | .029 | .029 | .000 |
| Portland | PDX | .028 | .022 | .006 | .075 | .056 | .019 |

1. - Since variance estimates are adjusted for degrees of freedom, it is possible for D to exceed the initial estimate of SB. In that event, W is defined as zero, and the revised estimate of SB is taken as D.

KEY: - SB = Unrevised estimate of combined background variance
D = Variance due to day of week alone
W = Remainder attributable to net seasonal effect of weather

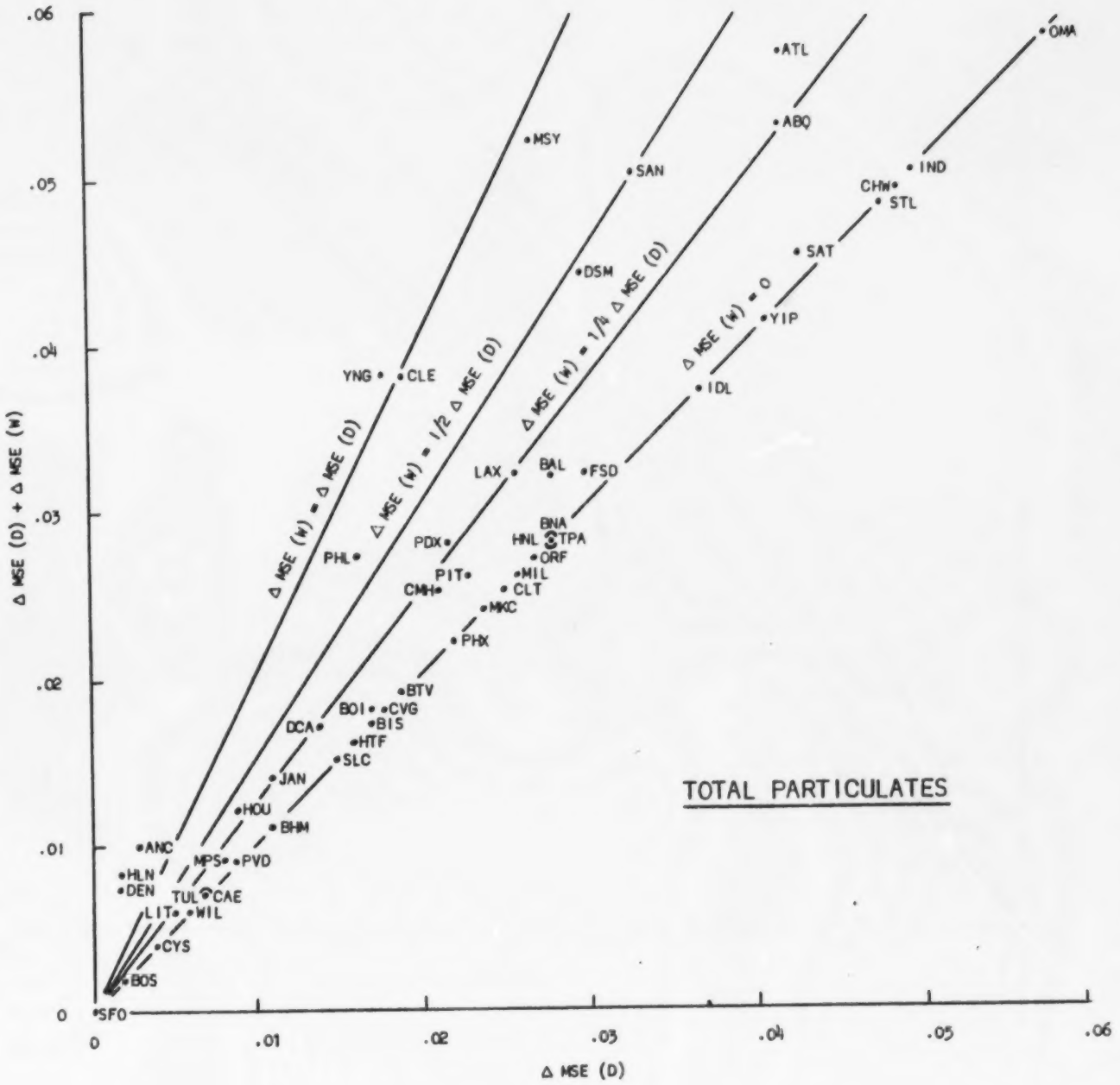



Fig. 14. The relative contributions of seasonal meteorological variability (ΔMSE (W)) and the day-of-the week (ΔMSE (D)) to the discrimination of HIGH and LOW total particulate pollution levels.

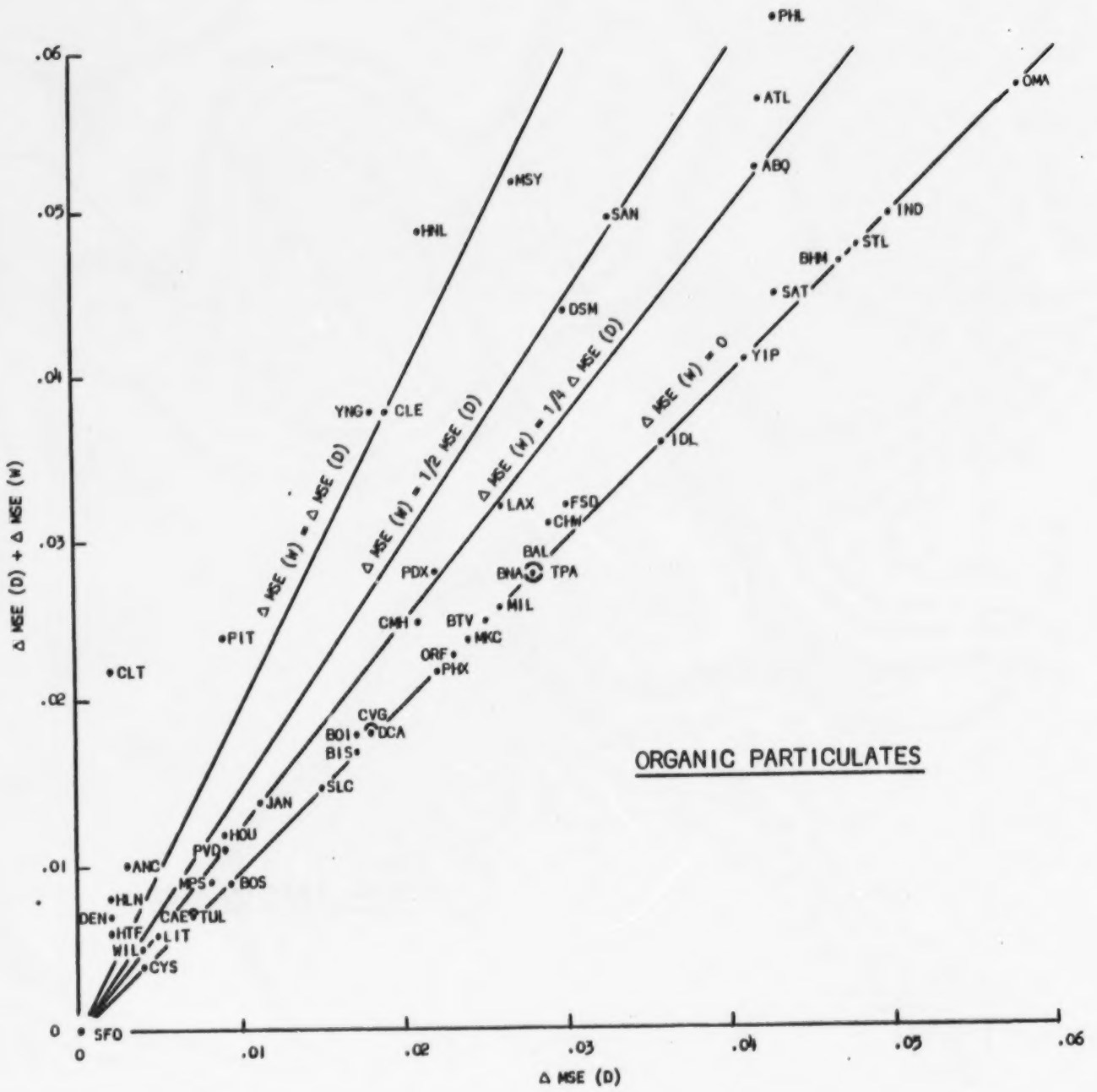



Fig. 15. The relative contributions of seasonal meteorological variability (ΔMSE (W)) and the day-of-the-week (ΔMSE (D)) to the discrimination of HIGH and LOW organic particulate pollution levels.

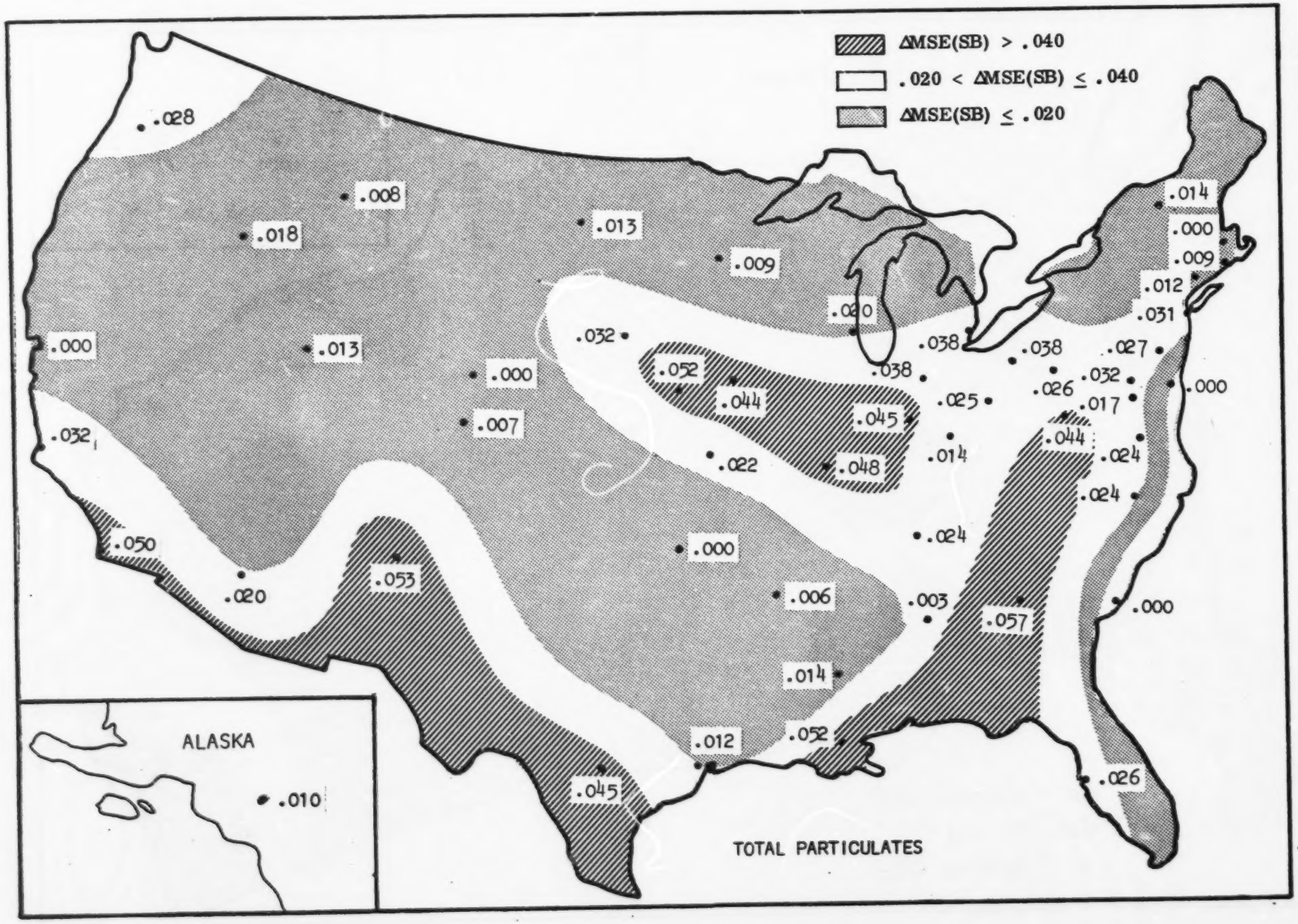



**Fig. 16.** The reduction of the mean square error of choice between HIGH and LOW pollution groups attributable to the seasonal weather and day-of-the-week variations, ΔMSE (SB.)

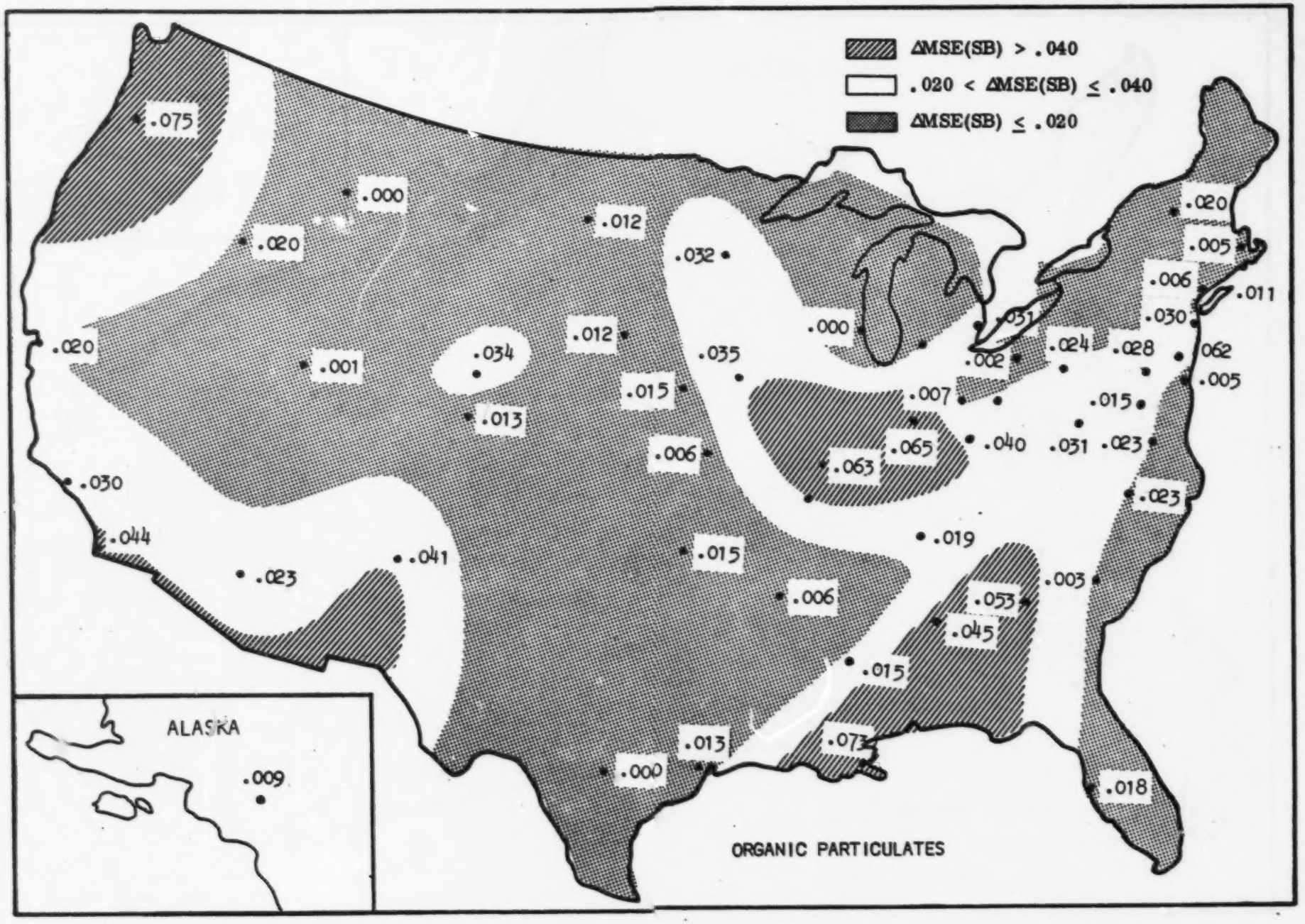



Fig. 17. The reduction of the mean square error of choice between HIGH and LOW pollution groups attributable to the seasonal weather and day-of-the-week variations, ΔMSE ($SI^3$.)

## C. The Contributions of Concurrent Meteorological Variability to Two-Group Pollution Prediction

Having disposed of the contributions of non-meteorological and seasonal weather variations, insofar as they have been measured and are discernible by the present data and analytic techniques, attention may now be turned to the primary purpose of this study, the determination of the predictability of two-group pollution levels from concurrent meteorological information derivable from routine U. S. Weather Bureau observations. The total contribution of these variables is discussed in this sub-section and a more detailed analysis of the relative contribution of each of the meteorological variables is presented in the next Section.

The total reduction in the mean square error of discrimination between HIGH and LOW pollution groups achieved by the concurrent meteorological variables has been summed for each station from the results presented in Table 24 of Volume I and plotted in Figures 18 and 19 of this Volume.

Study of the individual station values of $\Delta MSE(M)$ as shown in Figures 18 and 19 shows very substantial variations of the predictability of pollution groups from concurrent meteorological variables. The greatest predictability is for total particulates at Birmingham, Alabama, where $\Delta MSE(M) = .176$ or slightly more than 70 per cent of the variance of the occurrence of HIGH and LOW pollution levels is predictable from knowledge of the concurrent meteorological variables. At the other extreme, the minimum predictability for total particulates is found at Indianapolis, Indiana, and Atlanta, Georgia, where only slightly more than 10% of this variability is accounted for by conventional measurements of current weather conditions as measured at those cities.

The frequency of occurrence of various values of $\Delta MSE(M)$ is summarized in Table 4. As can be seen from the Table, current weather variables used in these studies account for 20 to 40% of the variance of occurrence

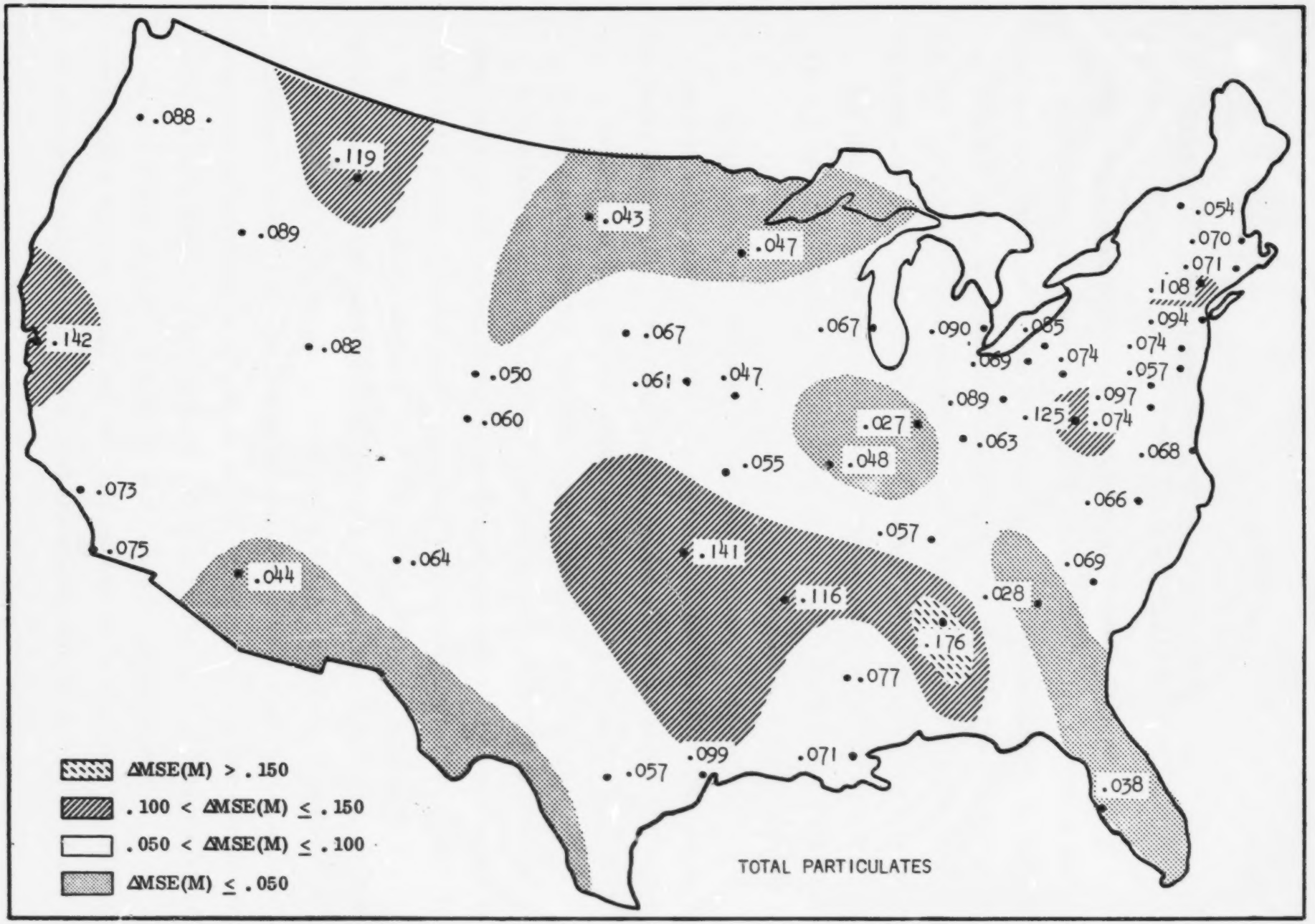



Fig. 18. The reduction of mean square error of discrimination of HIGH and LOW pollution levels attributable to concurrent meteorological variables (ΔMSE (M))

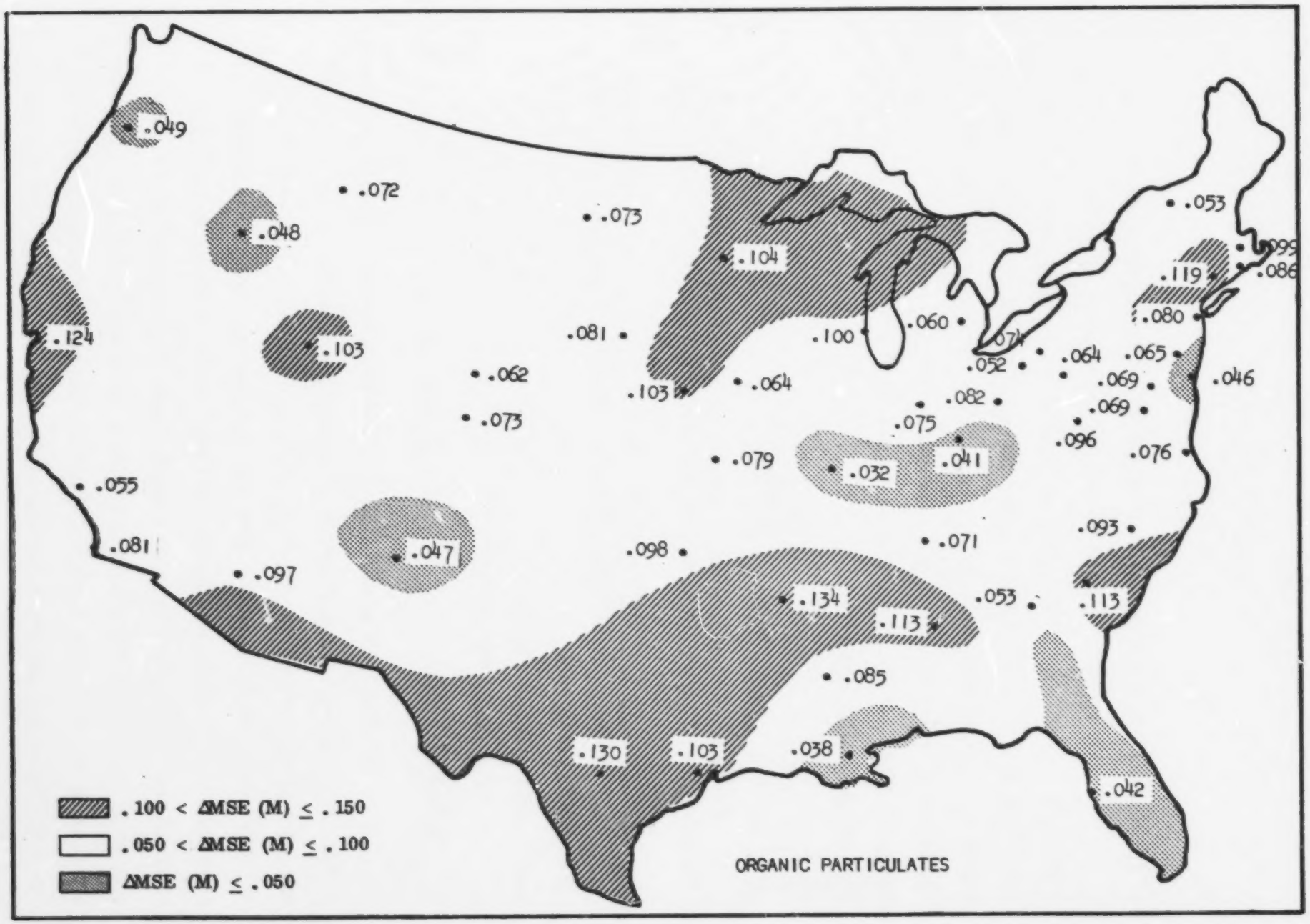



Fig. 19. The reduction of mean square error of discrimination of HIGH and LOW pollution levels attributable to concurrent meteorological variables (ΔMSE (M))

of HIGH and LOW pollution levels at a majority of the NASN stations.

Returning to Figures 18 and 19, the geographical distribution of $\Delta MSE(M)$ shows very little regional similitude for either total or organic particulates. With only minor exceptions, no systematic dependence of $\Delta MSE(M)$ on regional terrain, meteorological, or industrial-population criteria can be discerned. Departures of $\Delta MSE(M)$ from the highly favored range of .05 to .10 are, in most cases, local in that neighboring stations do not exhibit similar departures.

TABLE 4

Frequency of occurrence of various ranges of $\Delta MSE(M)$ at 49 National Air Sampling Network stations within the Continental United States

| Range of $\Delta MSE(M)$ | Total Particulates | Organic Particulates |
|---|---|---|
| $\Delta MSE(M) \leq .05$ | 7 | 8 |
| $.05 < \Delta MSE(M) \leq .10$ | 35 | 31 |
| $.10 < \Delta MSE(M) \leq .15$ | 6 | 10 |
| $\Delta MSE(M) > .15$ | 1 | 0 |

The levels of predictability of two-group pollution, as measured by $\Delta MSE(M)$, are highly significant at all except perhaps two or three stations, and point up the strong role of meteorological variables in the determination of local urban pollution levels. However, the lack of regional similitude also points up, with considerable force, the local nature of the urban pollution problem and the complexities of the local source-atmosphere-terrain interactions which produce the observed levels of atmospheric pollution. The residual mean square error, after $\Delta MSE(M)$ has been removed, must be attributed to source and terrain factors and to errors and inadequacies of measurement of the meteorological factors important in the context of urban air pollution,

all of which are presently unknown (except for the contribution of seasonal weather and weekly source cycles discussed previously). At approximately forty out of forty-nine cities these unknown factors contribute more than 60 per cent of the variance of occurrence of HIGH and LOW pollution levels. The need for more comprehensive measurements and further research into these complex problems appears obvious.

D. The Total Predictability of Two-Group Pollution Levels by Presently Available Independent Variables

The results of the analyses of predictability of HIGH and LOW pollution on the basis of day-of-the-week and seasonal and concurrent meteorological variables may now be combined for an appraisal of the predictability which can be achieved by application of two-group multiple discriminant analysis to the National Air Sampling Network data.

The total reduction in mean square error ($\Delta$MSE) achieved by the NASN predictors at each station has been plotted in Figures 20 and 21 and are summarized in frequency distribution form in Table 5. As can be seen from Table 5, the reduction in variance lies almost entirely in the 20% to 60% range.

TABLE 5

Frequency of Various Ranges of Total Reduction in Mean Square of Two-Group Pollution Level Discrimination Achieved By All NASN Predictors

| Range of MSE | Total Particulates | Organic Predictors |
|---|---|---|
| $\Delta$MSE $\leq$ .05 | 1 | 0 |
| .15 < $\Delta$MSE $\leq$ .10 | 24 | 25 |
| .10 < $\Delta$MSE $\leq$ .15 | 23 | 24 |
| $\Delta$MSE > .15 | 1 | 0 |

The minimum value of ΔMSE, also shown in Figure 13, is .05, a 20% reduction in variance, and the maximum value is .179, a 72% reduction in variance. Both of these extremes are associated with total particulate concentrations. All of the values of ΔMSE are highly significant at the 5% level.

With the reintroduction of the non-meteorological day-of-the-week predictor, which is distinctly regional in value, the geographical distribution of total predictability again assumes some more coherent regional definition. The general dominance of the meteorological variables over the day-of-the-week predictor is evident, however, in that the geographical pattern of ΔMSE(SB) (Figures 16 and 17) is not clearly evident in Figures 20 and 21.

Comparing Figures 20 and 21, the general geographic distributions of large and small predictability of total and organic pollution levels are quite similar except for the Minnesota - Wisconsin and Arizona - Utah areas where the relative predictability of total particulates is low but is high for organic particulates. The reverse situation is observed in Montana and Idaho. Areas of generally high predictability cover the West Coast, the Southwest, the South (with the exception of the Southeast) and the Middle Atlantic States including Ohio. Otherwise, with the exceptions noted above, the total predictability is relatively small for the remainder of the United States.

### E. Total Predictability as a Function of the Frequency Distribution of Concentration

Before turning attention to the relative contributions of the various concurrent meteorological predictors to the discrimination of HIGH and LOW pollution groups, it is worthwhile to examine the dependence of this discrimination on the frequency distribution of pollution at each station.

While the two-group discriminant analysis does not directly concern itself with the magnitude of the departure of concentration above or below the median value, it is reasonable to assume that large departures are asso-

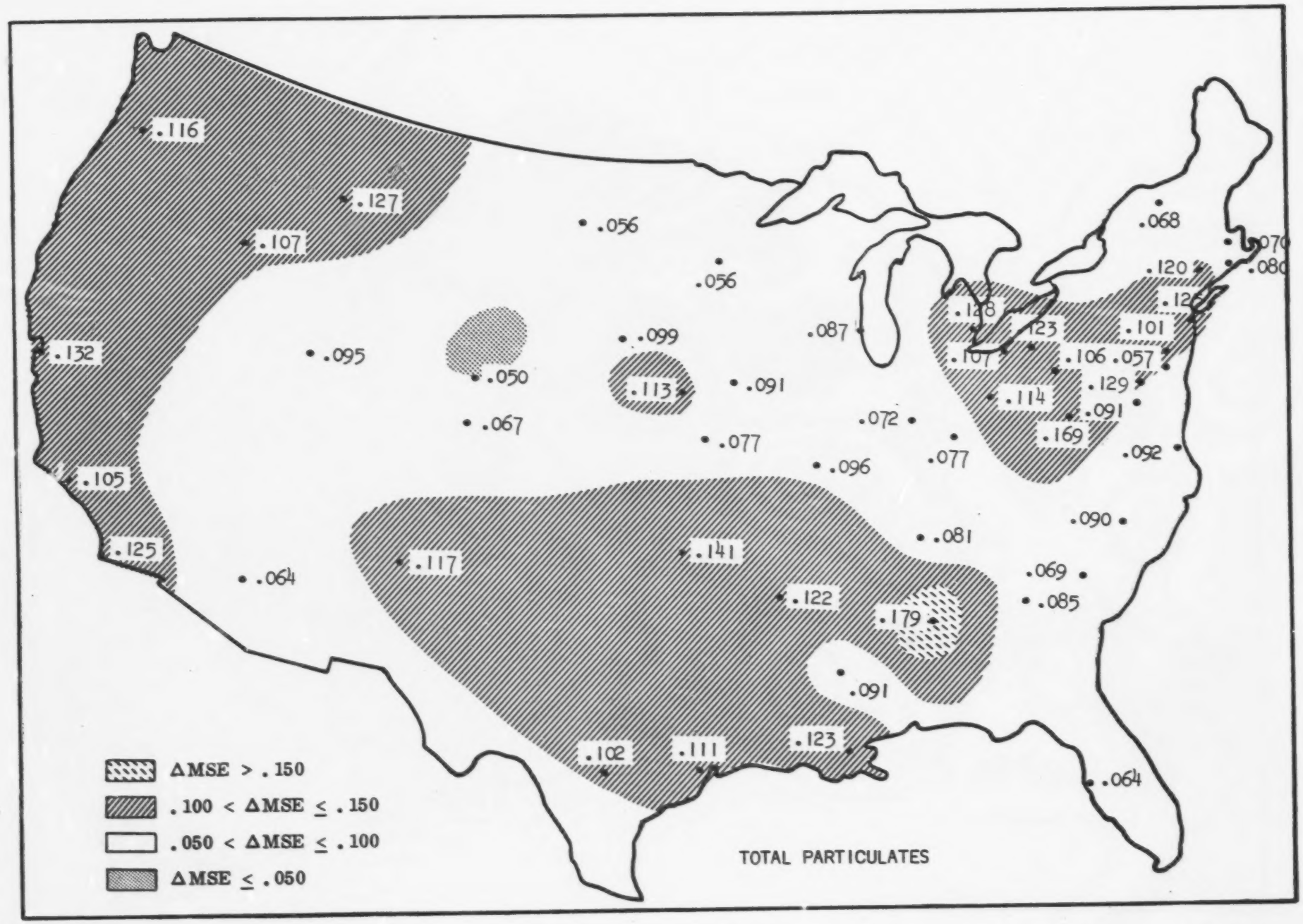



Fig. 20. Total reduction in mean square error of discrimination of HIGH and LOW pollution levels achieved by all predictors available from National Air Sampling Network Data.

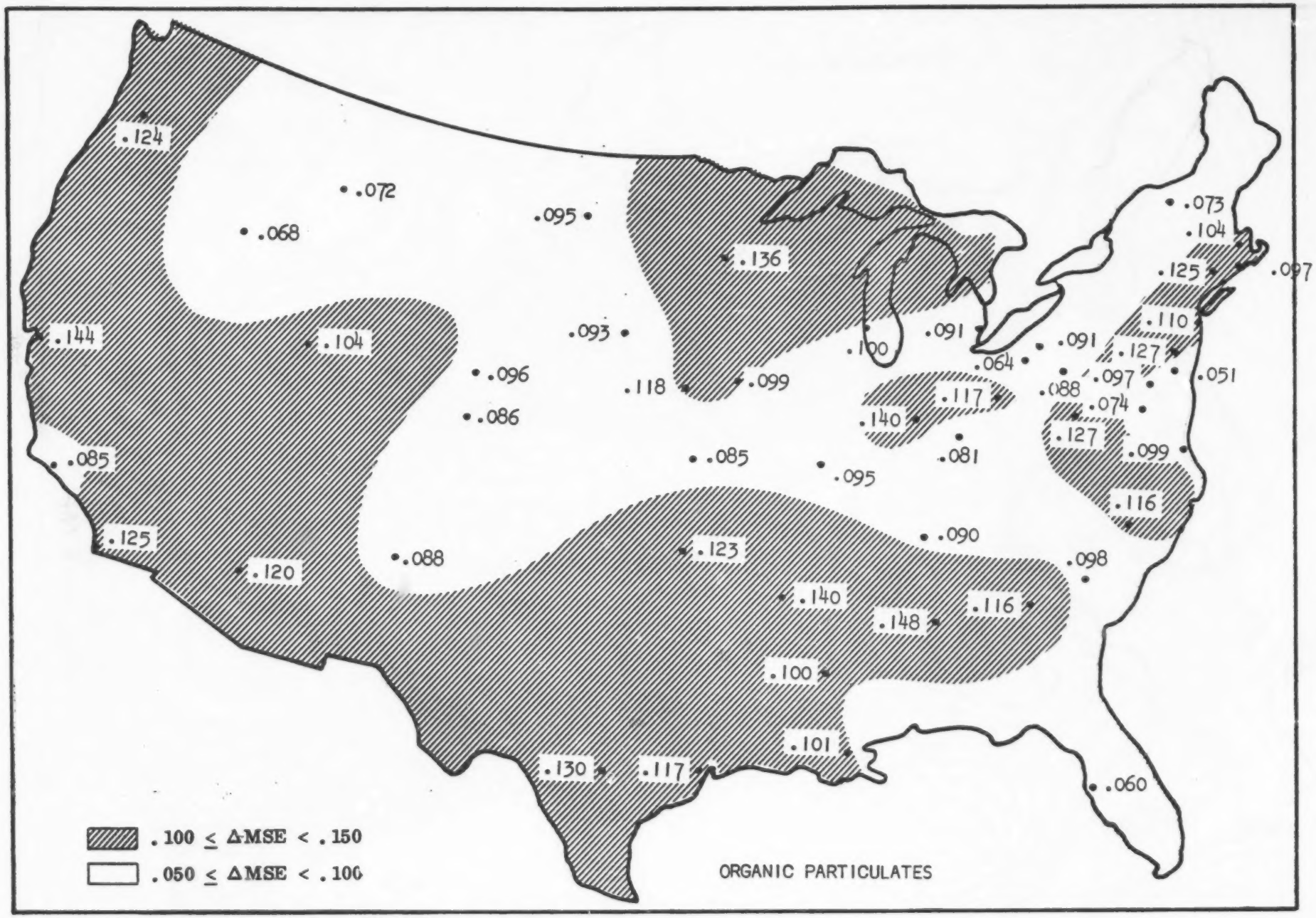



**Fig. 21.** Total reduction in mean square error of discrimination of HIGH and LOW pollution levels achieved by all predictors available from National Air Sampling Network Data.

ciated with comparably large variabilities in the meteorological parameters. If so, discrimination of only the sign of the departure (or group classification) should be better when the range of actual pollution values is large. The range of concentration experienced at any station is measured by the dispersion of the frequency distribution of this quantity, $\sigma_L$. On the basis of the above qualitative argument, $\Delta$MSE should increase with increasing $\sigma_L$. This will not be the case, of course, if $\sigma_L$ is not appreciably influenced by meteorological variability.

One can also argue that $\Delta$MSE may depend upon the general level of pollution, as measured by $\chi_{med}$ or $\bar{L}$. This argument depends upon the well-established fact that the range in concentrations of airborne pollutants due only to the variability of the dispersion capacity of the atmosphere is at least two orders of magnitude. (Averaging through a continuous 24-hour period tends to average out the variations of dispersion capacity associated with the diurnal atmospheric cycle, a prime contributor to these extremes.) If the total source of pollutants is small, the maximum concentrations associated with poor dispersion will also be small and the threshold of detection will be passed with only slight improvement of the dispersion. Then, with small source strength, there is only very limited sampling of the whole spectrum of atmospheric dispersion.

On this basis, prediction should be relatively poor when the median value of $\chi$ is very small and should improve with increasing median concentration.

These arguments may be tested against the NASN data by mapping the minimum mean square error, $MSE_{min}$ [1], in the $(\bar{L}, \sigma_L)$ space, as shown in Figures 22 and 23. Close examination of the distributions of $MSE_{min}$ contingent upon $\bar{L}$ and $\sigma_L$ shows a large degree of independence, with a barely (and somewhat subjectively) discernible axis of maximum predictability defined

---

1 $MSE_{min} = .250 -$ MSE.

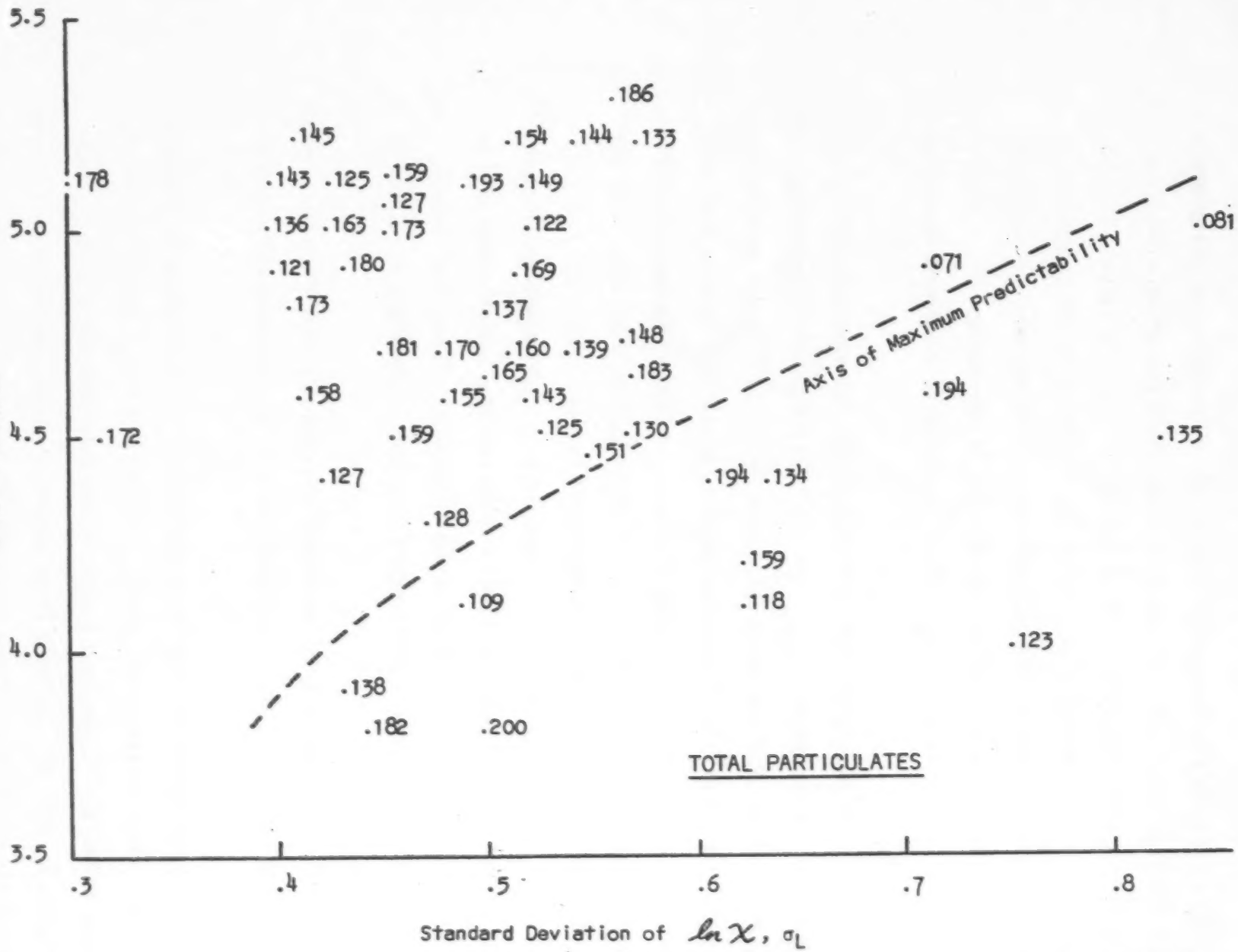



Fig. 22. Minimum MSE as related to central tendency and dispersion of frequency distributions of $\ln X$ for 51 urban stations. Data for period 1957 - 1959. (Individual stations may be identified by reference to Fig. 3.)

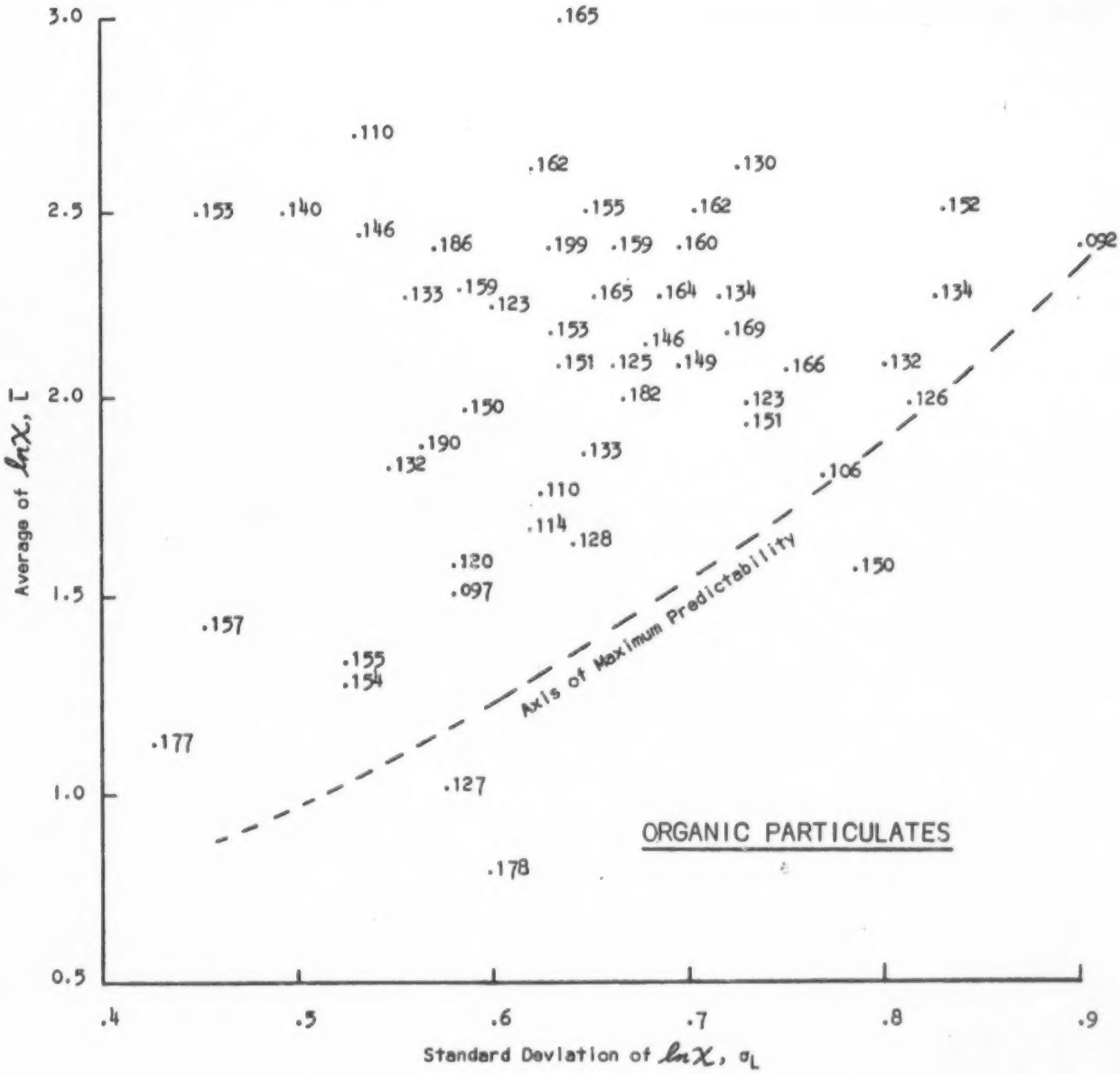



Fig. 23. Minimum MSE as related to central tendency and dispersion of frequency distribution of $\ln X$ for 51 urban stations. Data for period 1957 - 1959. (Individual stations may be identified by reference to Fig. 4.)

by the minimum of $MSE_{min}$ for each value of $\sigma_L$. The only truly significant feature of these diagrams is the association of the absolute maximum of predictability with joint maxima of $\bar{L}$ and $\sigma_L$. At lesser values of these parameters no clear dependence of predictability upon the range and level of pollution experience emerges.

## IV. COMPARISON OF THE PREDICTIVE INFORMATION CONTAINED IN THE NASN METEOROLOGICAL VARIABLES

So far only the total predictive capacity of the meteorological variables has been discussed. It is now of interest to view the separate effects of the meteorological variables. The individual predictors are listed in their rank order of choice along with the reduction in mean square error achieved by each predictor in turn in Table 24 of Volume I. These data constitute the basis for the following analyses.

As was seen in Figure 13, the reduction in MSE diminishes rapidly as the number of predictors increases. The analysis technique assures the linear independence of any predictor from any other predictor chosen, hence, this reduction in predictive capacity may be due either to redundancy of information content among the predictors or to real variability of the predictive information inherent in the various predictors. It must be noted, however, that the predictors selected and their rank order of importance does not constitute a unique set for the data analyzed by the discriminant analysis technique. Rather, it constitutes a "good" set which is probably close to the optimum, but which could be altered, because of possible marginal choices among predictors, upon the introduction of even small quantities of new data.[1]

The station-by-station results presented in Volume I lend themselves to a wide variety of further analyses and interpretation. It is not the purpose of this Report to go through a station-by-station analysis, however. Rather, a general comparison of the results for all stations and the definition of some of the more general implications of these results is attempted here.

### A. The Frequency of Selection of Each Meteorological Predictor and its Average Contribution to Two-Group Discrimination

Let $R$ be the rank order of predictor selection by the two-group discriminant analysis and $\Delta MSE(R)$ be the reduction in mean square error

1. For detailed discussion, see Volume I.

achieved by the Rth predictor. If N is the total number of times the predictor P is chosen in rank order R, as summed over the 51 stations, then the average reduction in mean square error attributable to P in the Rth predictor selection ranking is

$$\overline{\Delta MSE}\ (P,R) = \frac{1}{N} \sum_{N} \Delta MSE(P,R) \qquad (5)$$

The data on meteorological discrimination of HIGH and LOW pollution levels presented in Table 24 of Volume I are summarized according to the above parameters in Table 6. These summaries are compiled for the first five predictors selected, since in the vast majority of the cases studied, the contribution to $\Delta MSE(M)$ has become insignificant at this number of predictors

Even a casual inspection of Table 6 points up what is probably the most definitive result of these studies, namely, the clear dominance of wind velocity as the primary meteorological predictor. In the case of total particulates, some measure of the wind velocity (predictors 3, 10, and 11) is chosen first in 33 cases out of 51 and of these 33, wind direction is chosen first 28 times. In the case of organic particulates the wind direction is chosen first 16 times and a measure of the wind speed (predictors 3 and 11) is chosen first 32 times, for a total of 48 times out of 51 that a measure of the wind velocity is the first predictor! No other meteorological variable or class of variables even approaches this frequency of selection.

This dominance of wind velocity is not particularly surprising since, of the variables available, the wind speed and direction are the only direct measures of meteorological quantities known to be intimately related to the transport and dispersion of airborne materials. It is a clear commentary on the data and the analysis techniques that, without coercion, they identified this physically demonstrable relationship so clearly. In view of the fact that

TABLE 6

Summary of the frequency of choice of each meteorological variable and the average reduction in mean square error ($\Delta\overline{MSE}$), by predictor and order of selection, and by predictor only.

| Predictor (P) | Rank Order of Predictor Selection (R) | | | | | | | | | | Totals | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | | |
| | N | $\Delta\overline{MSE}$ | N | $\Delta\overline{MSE}$ | N | $\Delta\overline{MSE}$ | N | $\Delta\overline{MSE}$ | N | $\Delta\overline{MSE}$ | $N_T$ | $\overline{\Delta MSE}$ |
| | | | | | Total Particulates | | | | | | | |
| 3 | 2 | .054 | 4 | .021 | 4 | .006 | 1 | .001 | 3 | .003 | 14 | .016 |
| 4 | 6 | .050 | 6 | .015 | 4 | .004 | 2 | .002 | 2 | .001 | 20 | .026 |
| 5 | 2 | .055 | 5 | .024 | 6 | .007 | 3 | .004 | 3 | .005 | 19 | .016 |
| 6 | 0 | - | 0 | - | 4 | .006 | 2 | .001 | 4 | .004 | 10 | .004 |
| 8 | 0 | - | 3 | .019 | 2 | .001 | 3 | .002 | 1 | .002 | 9 | .007 |
| 9 | 3 | .046 | 6 | .018 | 5 | .007 | 4 | .006 | 1 | .001 | 19 | .016 |
| 10 | 28 | .052 | 9 | .016 | 4 | .008 | 3 | .007 | 1 | .002 | 45 | .037 |
| 11 | 3 | .050 | 6 | .017 | 5 | .007 | 3 | .002 | 4 | .002 | 20 | .014 |
| 12 | 3 | .037 | 3 | .013 | 9 | .006 | 6 | .004 | 0 | - | 22 | .011 |
| 13 | 4 | .031 | 6 | .017 | 4 | .006 | 4 | .006 | 2 | .005 | 20 | .014 |
| 14 | 0 | - | 3 | .015 | 1 | .002 | 3 | .005 | 5 | .002 | 12 | .006 |
| | | | | | Organic Particulates | | | | | | | |
| 3 | 18 | .059 | 6 | .018 | 4 | .008 | 2 | .001 | 2 | .004 | 32 | .038 |
| 4 | 0 | - | 4 | .026 | 4 | .004 | 0 | - | 2 | 0 | 10 | .011 |
| 5 | 0 | - | 6 | .012 | 3 | .005 | 2 | .001 | 1 | 0 | 12 | .007 |
| 6 | 0 | - | 0 | - | 5 | .007 | 4 | .002 | 1 | .003 | 10 | .005 |
| 8 | 0 | - | 1 | .022 | 2 | .006 | 1 | .003 | 3 | .001 | 7 | .006 |
| 9 | 2 | .035 | 6 | .016 | 2 | .004 | 4 | .002 | 0 | - | 14 | .013 |
| 10 | 16 | .051 | 16 | .025 | 7 | .005 | 4 | .004 | 0 | - | 43 | .028 |
| 11 | 14 | .064 | 2 | .006 | 5 | .002 | 5 | .003 | 1 | .003 | 27 | .035 |
| 12 | 0 | - | 2 | .008 | 4 | .004 | 5 | .003 | 7 | .001 | 18 | .003 |
| 13 | 1 | .022 | 4 | .014 | 7 | .008 | 2 | .002 | 2 | .002 | 16 | .008 |
| 14 | 0 | - | 4 | .016 | 9 | .003 | 5 | .003 | 2 | .002 | 20 | .006 |



124 333

the data are probably not completely representative and the analysis techniques are still crude and imperfect in light of their potential, this result should provide considerable hope and guidance for further work.

The relative emphasis placed on wind direction and wind speed in the cases of total and organic particulates is deserving of comment before proceeding to a discussion of the other variables. In the discrimination of HIGH and LOW total particulate pollution groups, wind speed is completely secondary to wind direction. (In fact, the temperature range, predictor 4, edges out wind speed.) However, in the case of organic particulates, wind speed is dominant, although wind direction remains a very potent runner-up. This relative dependence of total and organic particulate pollution levels on the components of wind velocity has also been noted[2] in analyses of extensive air pollution measurements made at Louisville. These observations may be reconciled by the following considerations of source distribution:

If the sampling station is located in the center of a uniform and steady area source, the wind direction will have no influence on the pollution level other than by way of a separate correlation with other meteorological variables which affect atmospheric dispersion. Only the wind speed component of velocity can have a direct affect on the pollution level. On the other hand, as sources are concentrated more and more into selected areas around the sampling site, the pollution level becomes more and more dependent upon the wind direction. The dependence of concentration upon wind speed remains, but may be masked by the large difference of pollution concentration occasioned simply by the wind being either from a source or non-source direction.

If this argument is valid, we may conclude that the sources of

---

2 - U.S. Public Health Service, "The Louisville Air Pollution Study", Government Printing Office (Chicago) (in press).

total particulates (primarily inorganic materials) are generally much more concentrated in urban areas than are the sources of organic particulates. This conclusion, arrived at quite independently of any direct knowledge of source distributions for these materials, appears to be confirmed by observation[1].

B. The Total Contribution of Each Predictor to the Reduction of Mean Square Error of Two-Group Discrimination.

The relative contribution of each of the meteorological predictors, including wind velcoity which has just been discussed, may be most conveniently summarized by summing, over all stations, the reduction of mean square error attributable to each of the first five meteorological predictors chosen. This is equivalent to forming the product $N_T \overline{\overline{\Delta MSE}}$ for the values listed in the last column of Table 6. When this is done and the predictors are arranged in the order of decreasing values of $N_T \overline{\overline{\Delta MSE}}$ the ranking shown in Table 7 results.

TABLE 7

Ranking of meteorological predictors on the basis of total contribution to the predictability of two-group pollution levels ($N_T \overline{\overline{\Delta MSE}}$). $f_+$ and $f_-$ indicate relative frequency with which the indicated variable was positively and negatively partially correlated with the pollution level.

| Rank Order | Total Particulates | | | | Organic Particulates | | | |
|---|---|---|---|---|---|---|---|---|
| | Predictor | $f_+$ | $f_-$ | $N_t \overline{\overline{\Delta MSE}}$ | Predictor | $f_+$ | $f_-$ | $N_t \overline{\overline{\Delta MSE}}$ |
| 1 | Wind Direction | - | - | 1.665 | Wind Speed * | .19 | .86 | 1.216 |
| 2 | Temperature Range | .94 | .06 | .520 | Wind Direction | - | - | 1.204 |
| 3 | Maximum Temp. | 1.00 | 0 | .304 | Wind Spd.-2 Group | - | - | .945 |
| 4 | Hrs. of Precip. | .14 | .86 | .304 | Hrs. of Precip. | .40 | .60 | .182 |
| 5 | Wind Spd.-2 Group | - | - | .280 | Cloudiness | .90 | .10 | .128 |
| 6 | Cloudiness | 1.00 | 0 | .280 | Wind Dir.Variability | .57 | .43 | .120 |
| 7 | Relative Humidity | .75 | .25 | .242 | Temperature Range | .91 | .09 | .110 |
| 8 | Wind Speed * | .44 | .56 | .224 | Maximum Temp. | .83 | .17 | .084 |
| 9 | Wind Dir.Variability | .62 | .38 | .072 | Relative Humidity | .94 | .06 | .054 |
| 10 | Amount of Precip. | - | - | .063 | Minimum Temp. | - | - | .050 |
| 11 | Minimum Temp. | - | - | .040 | Amount of Precip. | - | - | .042 |

* Measured by the cube root of the 24-hr. average wind speed.

1. Personel communication with R. A. McCormick and E. Tabor, R. A. Taft Sanitary Engineering Center, U. S. Public Health Service.

The total contribution of meteorological variables to the discrimination of two-group pollution levels among the 51 stations is emphasized by noting that the total mean square error associated with purely random choice of levels for 51 stations is 12.75 (Total MSE = .25 x 51 = 12.75) for each class of pollutant. The total reduction in MSE achieved by the meteorological predictors is 3.994 for total particulates and 3.135 for organic particulates, an average reduction of mean square error of between 20 and 30 per cent. This is a very significant contribution, pointing up the importance to meteorological variability in urban air pollution processes. At the same time, the residual mean square error identifies the need for improved meteorological information and the specification of pertinent non-meteorological variables.

In order to discuss the relative contribution of each of the meteorological variables to the predictability of HIGH and LOW pollution levels, it is very helpful to know the sign of the partial correlation between each variable and the pollution level. To this end, this information has been summarized in Table 7, also, in the form of relative frequency tallies of positive and negative partial correlation, where appropriate, for the first nine predictors as ranked in Table 7. (Wind direction has only positive or zero correlation; the two-group measure of wind speed ( predictor 11) shows a negative correlation in the sense that relatively high frequencies of wind speeds >7 mph are associated with relatively low pollution levels, but is otherwise ambiguous in the present context. )

The dominance of wind direction for total particulates and the overriding importance of wind velocity for organic particulates is again obvious. Beyond wind speed and direction, the other variables contribute somewhat differently for these two classes of pollutants. In both cases, Minimum Temperature and the Total Amount of Precipitation contribute the least discrimination, and can be dismissed as inconsequential. If $N_T \Delta\overline{\overline{MSE}} < .100$ is selected as the level of significance, Wind Direction Variability can

also be dismissed as unimportant in the discrimination of two-group pollution levels for total particulates. Similarly, Maximum Temperature and Relative Humidity are unimportant in the prediction of organic particulate concentrations. The remaining predictors may now be discussed in turn.

Precipitation

Although the total daily (24-hour) amount of precipitation is unimportant as a predictor, the Number of Hours of Precipitation figures prominently in both total and organic particulate concentrations. A similar finding has been reported by Dickson[1] who offers three possible explanations, only one of which appears appropriate to the present analysis. To quote Dickson[2], "...; 2) small amounts of precipitation are very effective in cleansing particulate matter from the atmosphere, making relatively unimportant any subsequent precipitation". This explanation needs physical verification, but appears quite reasonable as an interpretation of the prominence of Hours of Precipitation and the inconsequential predictive capacity of Amount of Precipitation.

Maximum Temperature and Temperature Range

While the contributions of these variables is marginal for organic pollutants, they figure relatively prominently in the discrimination of total particulates, ranking second and third in this category in Table 7. The reasons for this behavior are not at all obvious when the predominant sign of the partial correlation between these variables and pollution level is noted. Both are generally positively correlated with HIGH and LOW pollution frequencies, in terms of $f_+$ and $f_-$.

---

1 - Dickson, R.R., 1961. "Meteorological Factors Affecting Air Pollution of A City," BAMS, vol. 42, no. 8, pp 556-560.

2 - ibid, p 559.

A partial explanation for this relationship may be that large values of Temperature Range and anomalously high Maximum Temperature are associated with light wind speeds and clear skies, and this intercorrelation is imperfectly accounted for by the analysis technique. Light wind speeds are, of course, associated with higher concentrations at any time of the day. Clear skies are conducive to strong thermal mixing (and consequent low concentrations) during daylight hours, but they are also associated with strong radiation, stabilized temperature lapse rates, and high concentrations of pollutants at night. The preponderance of these effects may be as these results indicate.

It should be noted that Minimum Temperature was relegated to a very minor role partially because of the order in which Temperature Range, Maximum Temperature, and Minimum Temperature were considered in the analysis. In this context, temperature as such is also a partial source variable, being related, as it is, to heating and power consumption. None of these considerations offer a clear and unequivocal explanation of the dependence of total particulates upon temperature, however, and this result must be relegated to the category of physical verification and further research at the present time.

## Cloudiness

The partial dependence of pollution level upon Cloudiness emerges quite clearly in this analysis. HIGH pollution levels are associated with large values of sky cover and LOW concentrations with relatively clear skies. As shown by the relative value of $N_T \Delta\overline{\overline{MSE}}$, this effect is not large for either total particulates or organic particulates, but it is consistent among the stations for which this predictor was selected.

The possible effect of Cloudiness in the context of surface heating and cooling has been discussed above. Following that logic, it would appear the 24-hour average concentration for a day or large maximum and small

minimum concentration as determined by large values of surface heating and cooling is less than for a day of more steady concentrations occasioned by near-neutral lapse rate conditions associated with maximum cloud cover and minimum surface heating and cooling.

The present results must be viewed as inconclusive so far as clear delineation of physical cause and effect, associated with this complex interaction of heating and cooling and thermal and mechanical mixing is concerned. These effects are not large, relative to wind velocity, and their proper evaluation will require greater precision and sophistication in data and analysis techniques than has been brought to this study.

Relative Humidity

The final variable of any consequence, and this only for total particulates, is the relative humidity. This variable is generally positively correlated with pollution level. Two fairly obvious reasons for this correlation are; 1) the particulate materials adsorb or absorb water in the presence of high relative humidity and, due to their increased size or adhesive properties, they are collected more efficiently by the filters used to take these measurements; 2) high relative humidities are associated with the evaporation of surface water and relatively small vertical atmospheric diffusion rates. Viewed in the latter context, relative humidity is itself a secondary tracer, partially but significantly controlled by the same atmospheric exchange processes which control particulate pollution dispersion. The sources for moisture and particulate pollutants may be quite dissimilar, however. If the relative humidity is related to particle concentrations by virtue of wetting and consequent enhanced particle collection efficiency, the apparent dependence of concentration on Relative Humidity may very well be spurious.

C. In Summary

These composite results of the analysis of predictability of pollution

level purely from concurrent conventional meteorological measurements point up very strongly the role of wind velocity and offer interesting insight into the relative role of source distributions. Beyond this, some dependence upon temperature, cloudiness, precipitation and relative humidity is indicated, but this dependence is certainly secondary to wind velocity.

The relatively inconclusive delineation of the dependence of concentration upon secondary parameters, such as temperature and cloudiness, highlights the need for direct measurements of ventilation and vertical exchange rates in the surface layers over urban areas.

While some refinement of pollution prediction may arise from further studies of the kind reported here, based on more extensive data collections and more sophisticated analysis techniques, it is doubtful that the levels of predictability necessary for the kinds of urban pollution control programs envisioned by Stern[1] for the year 1985 can be achieved with these kinds of data.

---

1 - Stern, A., 1961. Transactions of the Symposium on Air over Cities, U.S. Public Health Service, R.A. Taft Sanitary Engineering Center, (in press).

## V. TESTS FOR HOMOGENEITY AMONG NASN STATIONS

The final objective of these studies has been to investigate the possibility of combining data from two or more of the National Air Sampling Network stations in order to increase the data sample for further statistical analyses. Secondarily, the criteria for such a combination of data would also serve to identify cities with similar pollution potential and make it possible to share among them their common experience and the efficacy of tested pollution control measures.

A three-fold basis for testing for similarities among cities has been chosen for the initial attack upon this problem. These are:

- Similarity of the distribution of pollution experience as measured by $\sigma_L$, the standard deviation of the logarithm of concentration.
- Identity of the first and second meteorological predictors chosen by the two-group discriminant analysis.
- Similarity of the reduction in mean square error of pollution level discrimination achieved by the first two meteorological predictors selected by the two-group discriminant analysis.

The second and third criteria emphasize, of course, the similarity of the dependence of local pollution levels upon local meteorological variables. The first criteria lumps all of the dependence upon source and terrain variables into the observed distribution of pollution levels in the presence of similar response to meteorological conditions. The general level of pollution, as measured by $\chi_{med}$ or $\bar{L}$ has not been included in these criteria since this quantity is primarily responsive to total source strength and would, therefore, tend to segregate cities which may be similar in all respects except size.

In applying these criteria it was also required that four or more stations meet the second criterion before a group could be established for comparison. This additional limit was set in order to minimize the inclusion of spuriously similar pairs, or trios, of stations.

Similarity of $\sigma_L$ for two stations can be tested rigorously by employing the "F" test for identity of two variances. This test depends only upon the degrees of freedom (number of observations less one) associated with each of the two variances. According to this test the values $\sigma_{L1}$, and $\sigma_{L2}$ are not significantly different if

$$\frac{1}{F(\nu_1, \nu_2)} \leq \left(\frac{\sigma_{L1}}{\sigma_{L2}}\right)^2 \leq F(\nu_2, \nu_1,)$$

where $\nu$ is the degrees of freedom and the subscripts indicate the respective stations.

In the present data $\nu$ ranges from 45 to 77 with a strong preference for about 65. The "F" test for $\nu_1 = \nu_2 = 65$ gives limits of

$$.82\ \sigma_2 \leq \sigma_1 \leq 1.21\ \sigma_2$$

at the 10 per cent confidence level. This suggests that if $\sigma_{L2}$ is within 20 per cent of $\sigma_{L1}$ the two may be considered identical at this level of confidence. The range of $\sigma_L$ is from .32 to .84 in the data to be considered.

While a rigorous test for determining that the reduction in mean square error is significantly different from zero is available, there is no obvious way of testing the significance of the difference between two independent values of $\Delta$MSE. The third criterion was, therefore, applied subjectively and only after the first and second criteria had been applied.

The procedure followed was to identify all stations with identical first and second meteorological predictors ($P_1$ and $P_2$) for each pollution type (total and organic). These groups were then tested internally for similarity of $\sigma_L$. The resultant sub-groups were then inspected for similarity of values of $\Delta$MSE.

The groupings which resulted are given in Table 8. While no claim can be advanced for an exclusive and exhaustive combination of stations on this basis, the results given in Table 8 show there are real similarities among stations in the National Air Sampling Network. Some of these combinations may be surprising initially. However, for the most part they appear quite reasonable.

Groups I and IV under organic particulates stand out as particularly strong and logical combinations of similar cities. The predictability or response to meteorological variables is high and the geographical situations for these stations are similar, at least insofar as they are all located in the Plains States or are close neighbors.

Various logics could be applied to the other groups derived from these considerations. However, the next step should be the pooling of data from the stations within a group and, after withholding some part of the data as an independent sample, reanalyze the pooled data and test the new results against the independent sample. This work is beyond the scope of the present study.

TABLE 8

Grouping of stations on the basis of similarity of first two meteorological predictors ($P_1, P_2$), standard deviation of logarithm of concentration ($\sigma_L$), and reduction of mean square error ($\Delta$MSE) associated with $P_1$ and $P_2$. N is the total number of observations used in the analysis of $\sigma_L$ and $\Delta$MSE at each station.

| Group | $P_1, P_2$ | Station | $\sigma_L$ | $\Delta MSE_1$ | $\Delta MSE_2$ | $\sum \Delta MSE$ | N |
|---|---|---|---|---|---|---|---|
| | | Total Particulates | | | | | |
| I | 10,9(1) | Cheyenne | .50 | .031 | .018 | .049 | 76 |
| | | San Diego | .53 | .038 | .013 | .051 | 63 |
| | | Denver | .57 | .037 | .011 | .048 | 73 |
| II | 10,11 | Cincinnati | .41 | .032. | .024 | .056 | 60 |
| | | St. Louis | .51 | ..033 | .012 | .045 | 70 |
| | | Youngstown | .45 | .064 | .012 | .076 | 75 |
| | | Detroit | .52 | .070 | .019 | .089 | 62 |
| | | Houston | .53 | .064 | .013 | .077 | 65 |
| III | 10,13 | Honolulu | .43 | .070 | .016 | .086 | 68 |
| | | Washington | .46 | .055 | .017 | .072 | 71 |
| | | Wilmington (?) | .49 | .041 | .016 | .057 | 68 |
| | | Organic Particulates | | | | | |
| I | 10,3 | San Antonio | .58 | .085 | .030 | .115 | 67 |
| | | Houston | .65 | .049 | .046 | .095 | 65 |
| | | Washington | .74 | .034 | .013 | .047 | 71 |
| | | Baltimore | .75 | .046 | .019 | .065 | 71 |
| II | 3,5 | Burlington | .52 | .046 | .001 | .047 | 71 |
| | | Cleveland | .57 | .048 | .009 | .057 | 75 |
| | | Albuquerque | .62 | .041 | .004 | .045 | 69 |
| III | 3,10 | Honolulu | .58 | .107 | .014 | .121 | 68 |
| | | Hartford | .60 | .102 | .007 | .109 | 76 |
| | | Youngstown | .66 | .049 | .022 | .071 | 75 |
| | | Des Moines | .73 | .049 | .008 | .057 | 77 |
| IV | 11,10 | Omaha | .55 | .076 | .024 | .100 | 68 |
| | | Tulsa | .58 | .082 | .020 | .102 | 54 |
| | | San Antonio | .58 | .085 | .030 | .115 | 67 |
| | | Milwaukee | .59 | .073 | .026 | .099 | 67 |
| | | Minneapolis | .63 | .079 | .024 | .103 | 70 |
| | | Providence | .63 | .059 | .013 | .072 | 74 |
| | | Kansas City | .64 | .057 | .018 | .075 | 68 |

1 - See Table 2 for identification of meteorological predictors.

# APPENDIX A

## Variables Reported in

## National Air Sampling Network Data

I. **Period of Record:** January 1957 through December 1959.

II. **Air Quality and Meteorological Data Used During Sampling Period:**

1. Total Particulate - $\mu g/m^3$
2. Organic - $\mu g/m^3$
3. Wind Direction - Frequency of Occurrence to 16 points.
4. Wind Speed - Frequency of Occurrence in the following classes:
   a. Less than 8 mph
   b. Greater than 7 mph
5. The Mean Wind Speed.
6. The Frequency of Occurrence of Relative Humidity in the following classes:
   a. 0-69 per cent
   b. 70-89 per cent
   c. 90-100 per cent
7. The Frequency of Occurrence of Total Cloud in the following classes in tenth of cloud cover:
   a. 0-5
   b. 6-9
   c. 10
8. The Maximum Temperature.
9. The Minimum Temperature.
10. The Total Amount of Precipitation.
11. The Number of Hours of Precipitation.